# Bloomberg Sports Visualization for Pitch Analysis

Bo Moon and Richard Brath

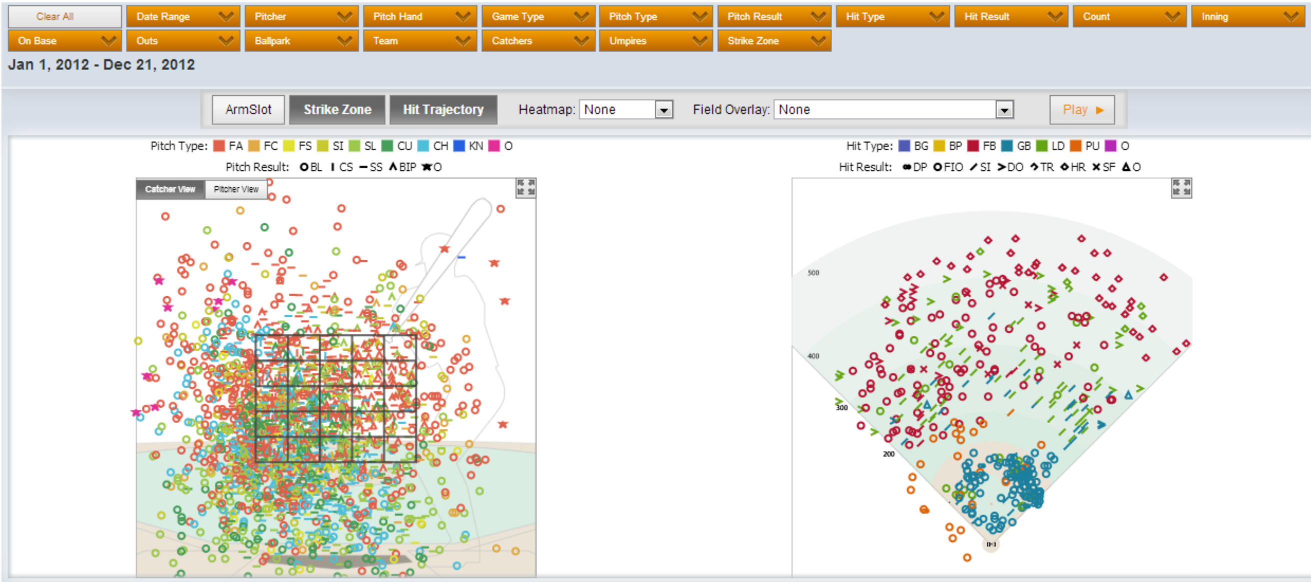


Fig. 1. The "Pitches" screen from the Bloomberg Sports professional application, showing all pitches to and hits by Josh Hamilton in 2012.

**Abstract**—Visualization of big data in baseball at the per pitch level including measured data and video can be made effective for coaches and managers through a variety of information visualization techniques focused on creating easy to use and easy to understand visualizations. Visualization techniques include filtering, linked familiar views, multi-variate glyphs with inituituve mappings to color and shape, aggregation via heatgrids and contour maps, and a workflow to support video analysis to and from the visualization.

**Index Terms**—Sports visualization, baseball pitch analysis, video analysis, big data visualization

## 1 Introduction

Data has been collected in baseball for more than one hundred years and deeply analysed such as The Book: Playing the Percentages in Baseball [TLD07] and popularized by the book and movie Moneyball [Lew03]. By the end of 2006, the volume of baseball data was expanding rapidly - data per each pitch was being collected, including:

- multiple quantitative data points per each pitch via the PITCHf/x system (a system that tracks speed and location of a pitched baseball from the pitcher's mound to home plate),
- analytical classification algorithms which determined what type of pitch was being thrown
- captured and tagged broadcast video for every pitch
- etc.

This results in multiple terabytes of heterogeneous data per year useful to a wide variety of different users – a "big data" situation.

Bloomberg Sports is a commercial company whose goal is to aggregate massive data and return value in unprecedented ways to both consumers and professionals in sports, leveraging technology built on the financial platform that leads global markets. One of Bloomberg Sports first products was an information platform for professional baseball combining a wide variety of data, video, news, analytic tools and insights for use by major league teams, broadcasters, players and agents. In early 2008, the organization determined there was an opportunity to leverage the new detailed per pitch data (i.e. PITCHf/x data, pitch classification and video) within this big data set. A number of data visualizations were designed and implemented in addition to other techniques. This paper discusses one specific visualization in the system called "Pitches".

## 2 Background

There are many precedents for data visualization in sports in general and baseball in particular, e.g. [Wil96, Fry05, CS06, Rei06, Tuf06]. With regards visual analysis of pitch and hit data, before the advent of PITCHf/x, there are a variety of examples, such as the traditional baseball scoresheet that can be used to record the type and location of a hit (http://en.wikipedia.org/wiki/Baseball_scorekeeping), Ted Williams strike zone chart from 1968 [Wil68] or more recently, the Baseball Fielding Bible (e.g. [Dew09]).

With the detailed data available from Sportvision's PITCHf/x in late 2006, the initial use cases included broadcast video (e.g. an inset graphic showing the pitch through the strike zone) and MLB's GameDay application, a 3D reconstruction of pitches for each player's "turn at bat" (e.g. http://mlb.mlb.com/mlb/gameday/y2009/). Quickly fans used the data to create a wide variety of visualizations to conduct analyses and post their insights and visualization images for web blogs such as fangraphs.com (e.g. [All09a]) and some

- *Bo Moon is with Bloomberg Sports LLC. E-mail: bmoon2@bloomberg.net.*
- *Richard Brath is with Oculus Info Inc.. E-mail: rbrath@oculusinfo.com.*

*Manuscript received 24 August 2013; accepted 20 Sept 2013.*
*For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.*

researchers began to investigate visualizations of this data (e.g. [Hun10]).

## 3 Pitches Visualization

Starting in 2008, a number of visualization designs and prototypes were created to view and analyse the variety of pitch level data. While the initial designs were wide ranging, reviews with stakeholders quickly narrowed in on a set of perceived requirements to be successful:

**Ease of understanding**. Representations needed to be easy to understand by target users. The visual layout, colors, icons, etc, all needed to be considered.

**Ease of use**. Representations needed to be easy to use. For example, a 3D view was discarded as it was considered more difficult to use, added little extra information, and made it more difficult to see minor variations in movement within the strikezone.

**Ease of analysis**. Simple analysis techniques, e.g. filtering on various criteria, needed to be easy to perform.

**Support for workflow**. A process to move from analysis of the quantitative data to the video data was required.

**Speed of response**. With an intended web application model, the application needed to respond fairly quickly with standard web technologies.

### 3.1 Linked Visualization

One of the key visualizations was a representation of the strike zone view showing each pitch over the plate. Around this core visualization, a number of "linked visualizations" were considered including a field view (showing each hit corresponding to the pitches); a pitcher view (showing the release point of each pitch); and various other views, such as timeseries of the pitch sequence, distribution of pitch speed, etc.

While "linked visualizations" can be powerful for analysis by updating all visualizations based on an interaction in any view, this linked visualization approach also has a number of problems that the second author encountered in earlier projects. Linked visualizations can:

- Appear overwhelming with many different visual panels all competing for visual attention,
- Increase compute requirements and slow down the user experience,
- Reduce screen real-estate for each view by needing to show many views.

As a result, the visualization (shown in figure 1) was ultimately simplified down to a default of two highly familiar views to the intended users – pitches through the strike zone and hits into the field. In addition, a panel of 17 filters via familiar drop down menus for attributes such as date range, pitch count and ball park were provided immediately above the visualization.

To perform a visual analysis the user starts with a specific player of interest – e.g. all pitches thrown to Josh Hamilton in 2012 are shown in figure 1. On the left side of the screen the strike zone is shown from the catcher's point of view displaying all the pitches as glyphs. Even though there may be thousands of pitches, some visual patterns may be immediately apparent, e.g. in figure 1 it can be seen there are more pitches toward the bottom left corner of the strike zone as opposed to the top right corner.

Similarly, on the right side of the screen a view of the field can be seen, showing all of Josh's hits in 2012. Again, some patterns may be immediately apparent, such as more deep hits in right field than left field.

The drop-down filters shown at the top can be used to filter data based on any combination of filter attributes, e.g. show only the pitches in the last month thrown to Josh in the last 3 innings of the game in a particular ball park.

The linked visualization analysis can be done by interacting with either view. For example, the user can click and drag around all the deep hits in the right view and all the corresponding pitches in the left view will be highlighted. Or all the pitches thrown through the top third of the strike zone can be selected to see corresponding hits. This interaction is useful to explore the relationship between the location of a pitch through the strike zone and the corresponding hits that have resulted.

### 3.2 Mnemonic Glyphs

Also with regards to ease of understanding, a variety contextual cues were included (e.g. wireframe image of a batter, field outline). For the point markers indicating pitches and hits, multi-variate glyphs were used building on past expertise (e.g. [Bra10, SB13]). The markers indicating pitches and hits are multi-variate glyphs indicating data attributes through color and shape using a connotative representation to aid the viewer in learning the representation and aid in decoding the representation. For pitches, the color hue indicated pitch type with faster pitches tending towards hot colors (figure 2), while the icon shape indicated the pitch result with simple geometric shapes representing five different outcomes with the representation based on a physical analogy (figure 3).

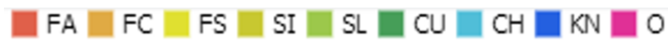


Fig. 2. Pitch type indicated by color, organized with fastest pitches in reds (fastballs), slowest pitches in blue (knuckleball) and breaking balls in greens and light blue. Pitch colors © 2008-2013 Bloomberg Sports LLC.

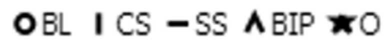


Fig. 3. Pitch result shown as connotative icons: circle for a ball, vertical line for a called strike, horizontal line for a swinging strike, chevron for hit into play, and star for other results (e.g. hit by pitch). Note that shapes use different visual shape features, such as curvature, horizontal, vertical, angular and corner attributes to aid visual discrimination when viewing a visual with many simultaneous glyphs. Pitch shapes © 2008-2013 Bloomberg Sports LLC.

Similarly, a different set of shapes and colors were used for hits, with color indicating the hit type and shape indicating the hit outcome (figures 4 and 5).


Fig. 4. Hit type indicated by color, organized with low hits in blue (BG bunt grounder, GB ground ball), high hits in red/orange/yellow (FB fly ball, PU popup, BP bunt popup), and green for line drives, purple for other results. Hit colors © 2008-2013 Bloomberg Sports LLC.

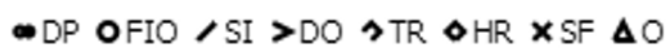


Fig. 5. Hit result shown as mnemonic icons, left to right:
- double play (an infinity symbol - each O indicating an out),
- field out (single O, larger than each loop of the double play),
- on-base hits single, double, triple and home run (each shown as a line indicating the runner's path from a top view)
- sacrifice fly (an X or cross)
- other (triangle)

Note that the size of the shapes are fairly consistent so that no shape is more dominant in the display than others. – e.g. the side of the shape single is larger than the side of the homerun . Hit shapes © 2008-2013 Bloomberg Sports LLC.

While both hue and shape are used for both pitches and hits, there is a different palette of hues and different palette of shapes for each, in order to create a unique set for each visualization: as users gained more familiarity with the system, they would immediately know which visualization type they were looking at (hit or pitch) even if their visual attention was focused on a few icons as opposed to the broader contextual cues across the visualizations.

### 3.3 Pitch Quantity and Summarization

One early concern was the response of the visualization when displaying a large number of pitches. With regards to displaying individual pitches as glyphs, effort was made to optimize performance with large numbers of pitch glyphs as well as features to enhance the display, such as automatically adjusting the point transparency based on the number of visible pitch icons in the display.

One desired feature was a summarization of pitches when displaying a large number of pitches. A contour map with a heat-based color scheme was the preferred approach and example images of such contour heatmaps were available on the web [All09a, All09b].

In 2008, a quick contour heatmap was constructed but was deemed to perform too slowly to be effective in the target hardware and target deployment platform (Adobe's Flash was the target platform 2008-2010). Given tight timelines to launch the software, instead a simple heatgrid was implemented (figure 6 left). While this was considered a reasonable "stop gap" the desire for a contour heatmap was still expressed by some users. The heatgrids have since been replaced by a contour heatmap when the site moved to HTML5 technologies (figure 6 right).

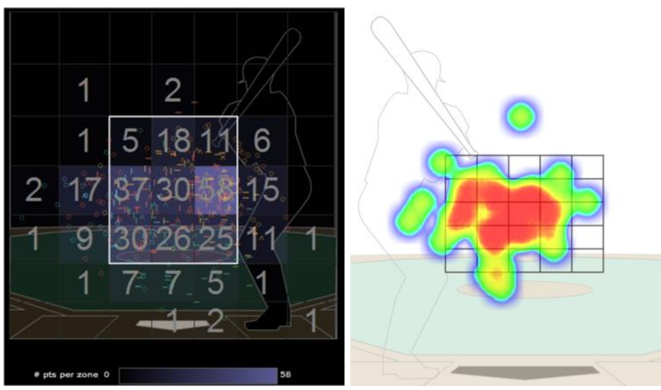


Fig. 6. Initial heatgrid implementation from 2009 (left) and revised heatmap implementation (right) from 2011.Users preferred the rainbow-colored contour heatmap.

### 3.4 Video

For most of the pitches there are video clips available. As of writing there are more than 2.5M video segments available and increasing. Video provides a wide variety of soft information not otherwise visible in the quantitative data, such as body position, mechanics of the pitch delivery, visible stress, etc. While video analysis has been used in baseball for decades, the opportunity is to provide an easy means to search and find the video corresponding to pitches of interest.

While a user focusing on the visualization alone for visual analysis tasks may narrow down from 1000's of pitches down to tens or hundreds; for the task of video analysis, the user may want narrow down to just a few pitches.

The solution uses the same visualization tools to narrow in on the pitches of interest through the filtering of any combination of the 17 attributes (e.g. date range, pitch type, pitch speed, etc) in combination with direct click-and-drag selection in any of the visualizations. This enables the user to quickly isolate pitches of interest, such as "All of Prince Fielder's deep hits off breaking balls to the outside of the strike zone" or "All of Andy Pettitte's cutters to right-handed batters in the last two months resulting in hits."

Based on the filters and selections, the users could transition to a video playback environment with a queue of video clips, detailed data per clip including play-by-play description, and various video controls such as slow-motion playback.

## 4 Results

Given the constraints of rapidly assembling a commercial product, requirements and feature sets evolved over the course of the work. In addition, limited resources meant that a best effort had to be pursued, with rapid feedback and ad hoc evaluation by experts revising functionality as work progressed.

The visualizations were part of a much broader application and was quickly adopted by a few MLB teams. Feedback from the teams indicated a strong interest in the visualizations by pitching and batting coaches as well as some players. Overall, response has been strong and currently more than 70% of MLB teams use Bloomberg Sports professional baseball application.

When the fan community found out about the visualizations in the pro product, blog reviewers posted comments such as: "Where the product really excels, however, is in the pro version. David Appelman has photos from the presentation, and the spray charts, strike zone info and PITCHf/x analysis are where this tool emerges as something drool-worthy." [Kab10]

One interesting point is that the use of the visualizations and videos together was originally thought to be a linear workflow (start with lots of pitches, narrow down, view video) but some users follow a much more fluid process moving back and forth between video and pitch visualizations (e.g. [Pet11]) – this enables comparison between different sets of pitches of interest, e.g. over different time periods to see and evaluate changes in pitch mechanics over time.

With easy to use point and click tools, new analytical questions can be hypothesized and answered. Within the project team, for example, people have investigated ad hoc questions that may have been more difficult to do previously, such as, "how do top hitters compare with hits relative to the strike zone?" [SB12], or "what's the relationship between umpires and catchers on called strikes that seem to be outside the zone?" or "what are the characteristics of hits in clutch situation?"

Another interesting point is that while the visualizations were effective for analytical uses such as pitching and hitting analysis, they were not necessarily useful as-is for other professional uses, (e.g. broadcast video, scouting) and additional adaptations were required.

Since creating this pitch visualization in 2009, this type of visualization of the strike zone and/or field has become more common in other uses too, including ongoing visualizations by fans (e.g. fastballs.wordpress.com), in broadcast network graphics, and in the popular press (e.g. [RCW10]).

## 5 Conclusions and Future Work

Bloomberg Sports visualizations are relevant to the visualization community as these visualizations show how effective visual analysis by domain experts (in many cases these users do not have much computer expertise) can be successfully accomplished in a big data set of mixed numeric, categorical and video data by focusing on creating an appropriate visualization user experience. By focusing on ease of understanding, use and analysis in addition to workflow and responsiveness visualization techniques such as filtering, linked familiar views, multi-variate glyphs, aggregations and video analysis have been deployed to many domain experts.

The use of visualizations in the Bloomberg Sports product have been successful, although there have been a number of zigs and zags in the product development such as completely changing the technology platform and evolving the visualization feature sets over different versions to meet the needs of different types of users such as the needs of broadcasters vs. the needs of coaches.

Bloomberg Sports success in baseball has led to the diversification of their baseball offerings as well as provide solutions in other sports. Bloomberg Sports has shown that a combination of data and analytic expertise, visualization design and development capabilities and domain expertise can create effective solutions to complement a sports analysis product. There are many areas where these visualizations could be extended within in baseball (e.g.

automated display of a drift or shift over time displayed visually as a path; or additional data coming from new emerging sources such as FIELDf/x data or biomechanics data); or extended to sports where there is a constant flow (e.g. basketball, soccer, hockey); or adapted to sports where there is a lot of activity within each event (e.g. per each down in NFL football).

Furthermore, some of these techniques and capabilities emerged from decades of expertise in financial services. From a broad viewpoint, the techniques presented here – visualization and video review in big data environment – can be applied to other transaction oriented processes with video capture, such as retail sales and merchandising, casino floor analytics, and video surveillance analysis. These different environments would require additional adaptations ranging from data integration through to potentially new visual representations.

## REFERENCES


[All09a] D. Allen. "45 Up, 45 Down." *FanGraphs.com* www.fangraphs.com/blogs/45-up-45-down/ July 29, 2009.

[All09b] D. Allen. "One of the Game's Stranger Hitters" *BaseballAnalysts.com* baseballanalysts.com/archives/2009/08/one_of_the_game.php. August 14, 2009.

[Bra10] R. Brath. Multiple Shape Attributes in Information Visualization: Guidance from Prior Art and Experiments. IEEE Information Visualization 2010 (IV10).

[CS06] A. Cox, J. Stasko. Sportsvis: Discovering meaning in sports statistics through information visualization. Compendium of Symposium on Information Visualization, 2006

[Dew09] J. Dewan. *The Fielding Bible Volume II*. ACTA Publications. Skokie IL, 2009.

[Fry05] B. Fry. Ben. "Salary vs. Performance." http://benfry.com/salaryper/2005/ 2005.

[Hun10] J. Hunt. Baseball Visualization: A Major Qualifying Project Report. Worcester Polytechnic Institute Electronic Projects Collection. 2010.

[Kab10] B. Kabak. The Visual Analytical Tools of Bloomberg Sports. http://riveraveblues.com/2010/02/the-visual-analytical-tools-of-bloomberg-sports-23238/ Feb 2010.

[Lew03] M. Lewis. Moneyball. Norton & company, NY, NY. 2003.

[Pet11] R. Peterson. Bloomberg Sports Pro Product: Rick Peterson on John Axford. www.youtube.com/watch?v=uUYY1kYrR8g Feb 13, 2011.

[RCW10] G. Roberts, S. Carter, J. Warden. How Mariano Rivera Dominates Hitters. NY Times Magazine. 2010. http://www.nytimes.com/interactive/2010/06/29/magazine/rivera-pitches.html?_r=0

[Rei06] A. Reisner. What's the Matter with Chernoff Faces? http://www.alexreisner.com/baseball/chernoff-faces 2006.

[SB12] N. Schwartz, R. Brath. Making Major League Data Work. Strata + Hadoop World. 2012. slides 30-36.

[SB13] E. Sorenson, R. Brath. Financial Visualization Case Study: Correlating Financial Timeseries and Discrete Events to Support Investment Decisions. IEEE Information Visualization 2013 (IV13).

[TLD07] T. Tango, M. Lichtman, A. Dolphin. The Book: Playing the Percentages in Baseball. Potomac Books, Dulles, VA, 2007.

[Tuf06] E. Tufte. Beautiful Evidence, Cheshire, CT: Graphics Press, 2006.

[Wil68] T. Williams. Strike zone chart tedwilliams.com/_data/hzone.htm

[Wil96] G. Wills. Selection: 524,288 ways to say this is interesting. In Proceedings IEEE Information Visualization 96. 1996.