

PSYC6135 PSYCHOLOGY OF DATA VISUALIZATION

VISUALIZATION OF BASEBALL PITCHER DATA



“From the beginnings of American professional baseball in 1869, statistics such as runs scored, batting average, and counts of home runs were used to measure accomplishments of ball players”
 (Jim Albert, 2018, p vii)



BASE BALL WITH CARDS.
The greatest GAME in the world.

A ROUND Table Game for old and young. A perfect game of Base Ball with Cards, and the greatest invention of the age being the first improvement in over 2000 years in playing cards.

PRICE, 50 CENTS PER PACK.

Sample pack sent to any address, postpaid, on receipt of the price. Manufactured by the

LAWSON CARD CO.,
C. A. BURGESS, Manager,
10 BEDFORD STREET, BOSTON, MASS.
For sale by all Stationers, Newsmen and Fancy Goods Dealers. Liberal discount to the trade.

LAWSON'S PATENT GAME.

BOSTONS.	1	2	3	4	5	6	7	8	9	10	AM.	HR.	P.	A.	E.	R.
SUTTON, 3 b.																
PODDEMAN, r. f.																
WISE, 2 b.																
MERRILL, 1 b.																
WHITNEY, p.																
MANNING, c. f.																
GUNNING, c.																
HACKETT, s. s.																
I. f.																
Umpire,																
Total,																
Game Began,																
I. B. Hit,																
L. on B.																
Struck out,																
First Base,																

The most complete, authentic, interesting, newsy and official base ball news is contained in the columns of the **BOSTON HERALD**. It is the only daily that publishes the official scores of Boston League Games, and its base ball writers are the most experienced and most competent in New England. If you want to be up to the times concerning all matters, read the **HERALD**.

COLUMBIA
THE POPULAR STEEDS
OF TODAY
BICYCLES

COLUMBIA TRICYCLES

FOR LADIES
AND
GENTLEMEN

ILLUSTRATED CATALOGUE

THE POPE MFG. CO.
317 WASHINGTON STREET
BOSTON
15 NASSAU ST. N. Y. CHICAGO ILL. PHOENIX

Dame, Stoddard & Kendall,
SUCCESSORS TO
HREADFORD & ANTHONY,
HEADQUARTERS FOR
FINE FISHING TACKLE
Of Every Description.
374 WASHINGTON STREET,
Opposite Bromfield Street,
BOSTON.

CHICAGOS.	1	2	3	4	5	6	7	8	9	10	AM.	HR.	P.	A.	E.	R.
DALRYMPLE, I. f.																
GORE, c. f.																
KELLY, r. f.																
ANSON, 1 b.																
PEPPER, 2 b.																
WILLIAMSON, 3 b.																
BURNS, s. s.																
CLARKSON, p.																
FLEET, c.																
Umpire,																
Total,																
Game Began,																
I. B. Hit,																
L. on B.																
Struck out,																
First Base,																

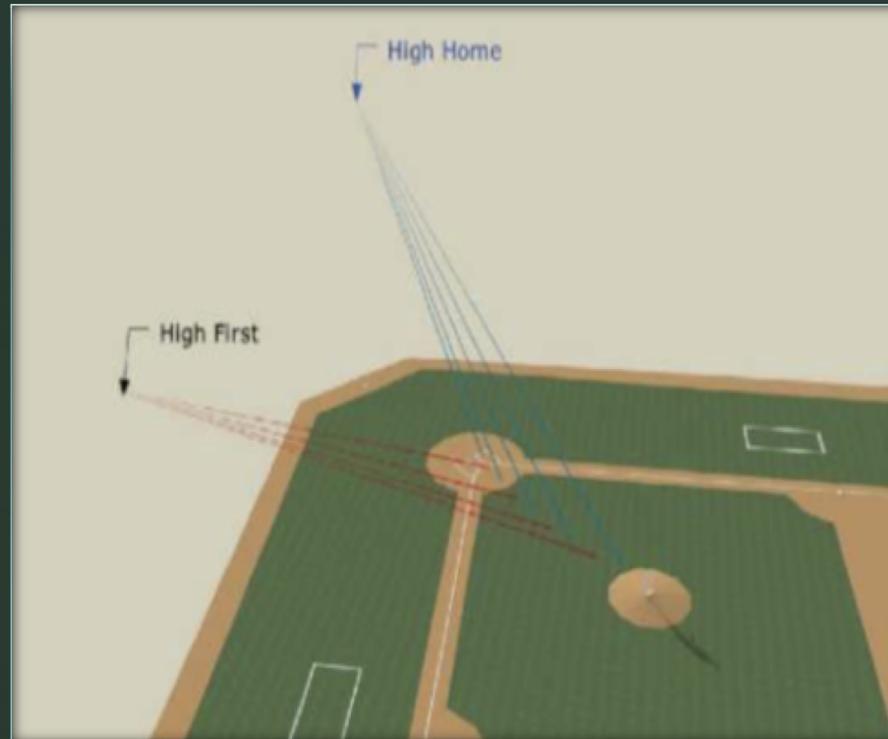
P. LORILLARD & CO'S PLUG TOBACCO.

SMOKE **DEFIANCE**
AND
CHEW **CLIMAX.**

“The language of information transmission is laden with traps that lead us away from the concerns of performance....” (Wurman, 2000, pp. 55)

▶ PITCH f/x Overview

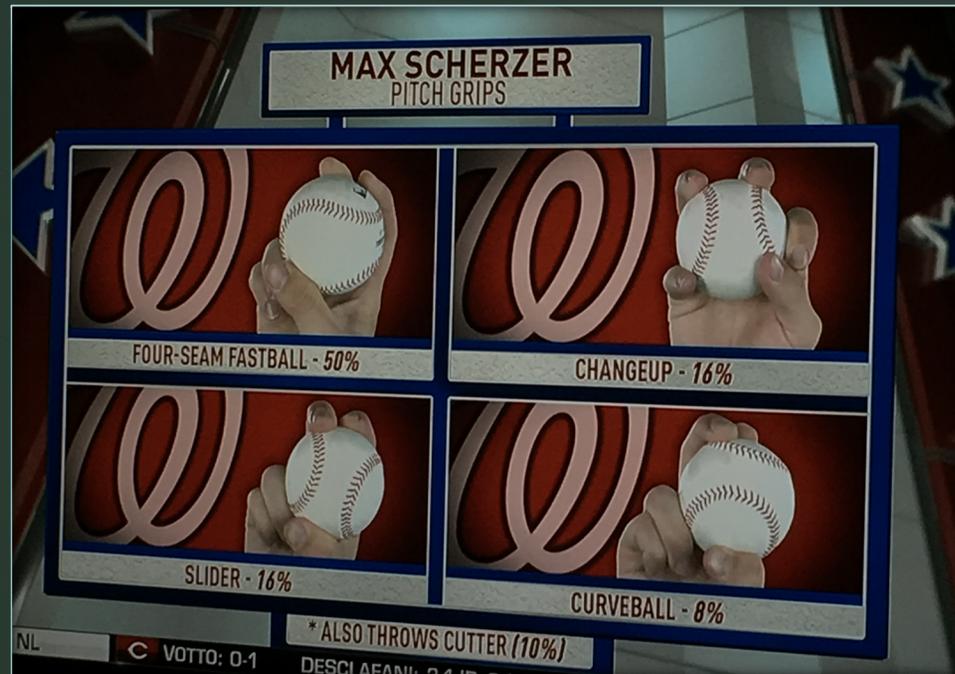
- Created by Sportvision & installed in stadiums as early as the end of the 2006 season
- Camera based system that tracks pitches in real time
- Calculates movement, speed, location and batting outcome at any time during the time of the pitch
 - Small margin of error (0.4 inches)



In some ways, PITCHf/x is a bridge between scouting and analysis, giving us an objective window into the batter-pitcher matchup at a level we've never seen before (Fast, 2010).

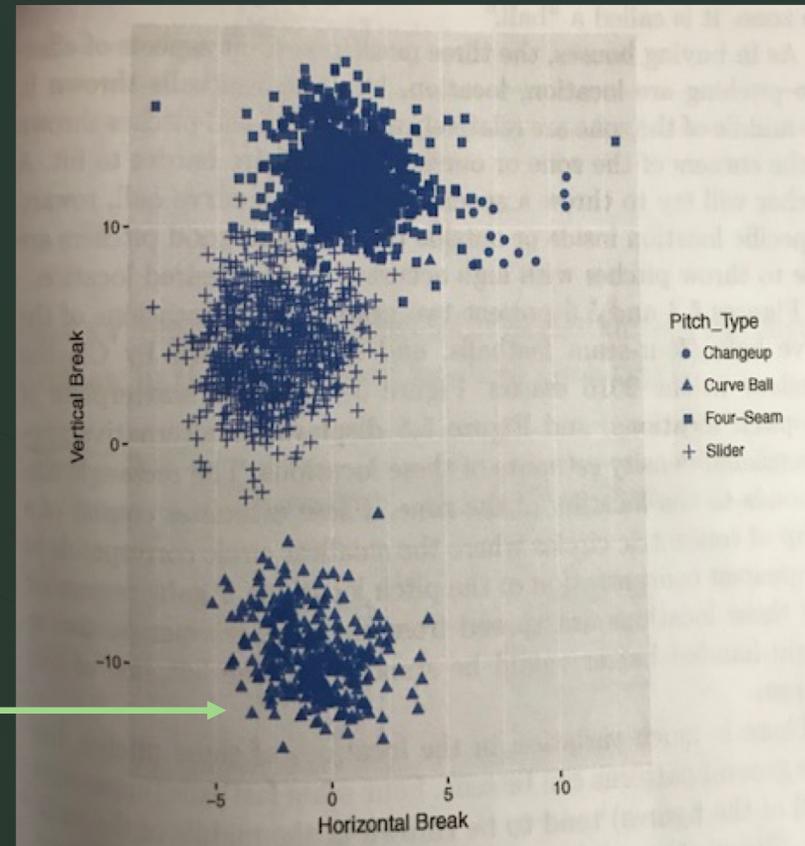
What questions can data visualizations with PITCH f/x answer?

- What do pitchers throw and when do they throw it?
- What are their most/least effective pitches?
- What makes an effective fastball, curveball, slider, ...?
- •speed, break, location, ...
- How do hitters perform against different pitch types, locations, speeds, etc.?
- What is effect of pitch sequencing?



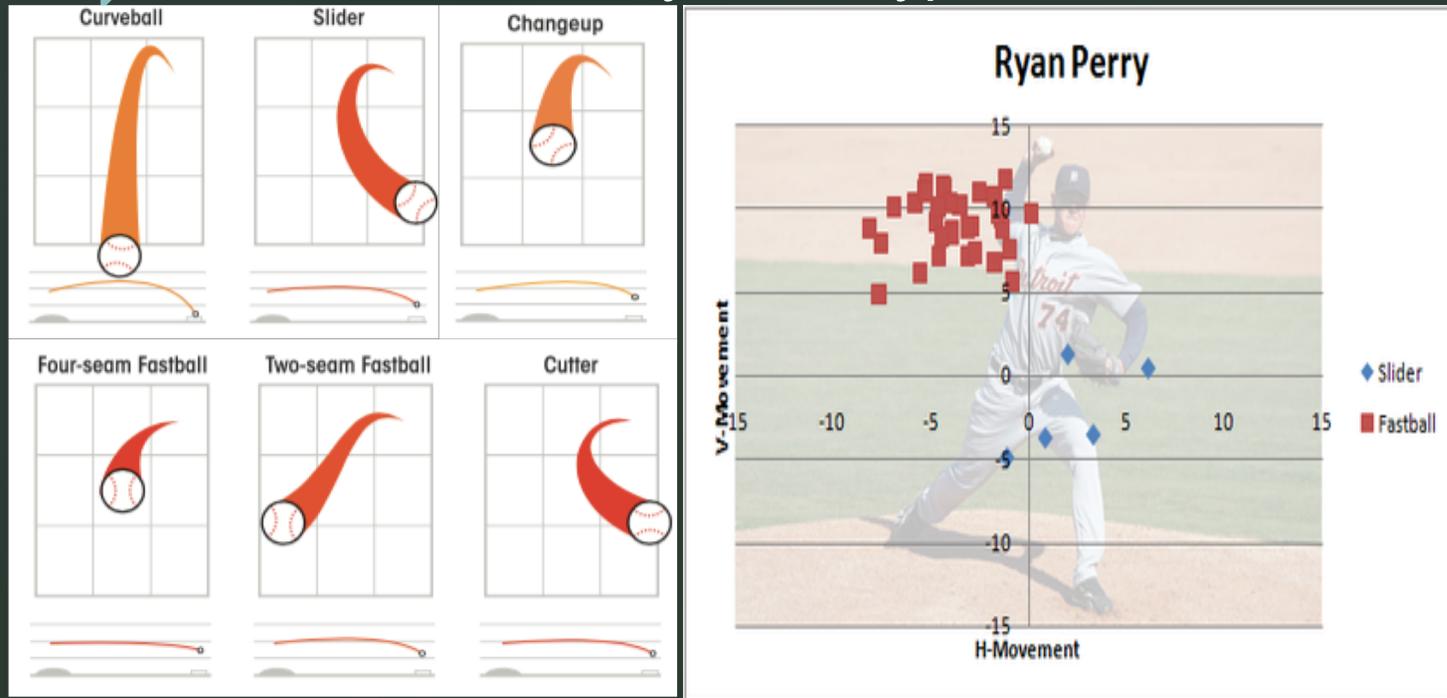
PITCH f/x Classify Pitch Type: Movement

- Measurements are made on the velocity and movement in the horizontal and vertical directions, and these measurements are used to classify each pitch as specific type” (Albert, 2018, p56).
- Aspects of a type of pitch for a specific pitcher.
 - Example:** What’s depicted here?
 - Movement of a curve ball is the strong negative vertical movement that drops the pitch below the lower edge of the strike zone.



Helpful for batters and batter coaches trying to predict the pitch

PITCH f/x Classify Pitch Type: Movement



“You could begin to build models that would help you predict a certain player at a certain age that plays a certain position that has had this type of résumé in the minor leagues and in his early major-league career. This was roughly what we would project him to do next year and the year after.” (Quarterly Review, July 2018)

➤ PITCH f/x Classify Pitch Type: Speed

impression of variability in one clean graphic

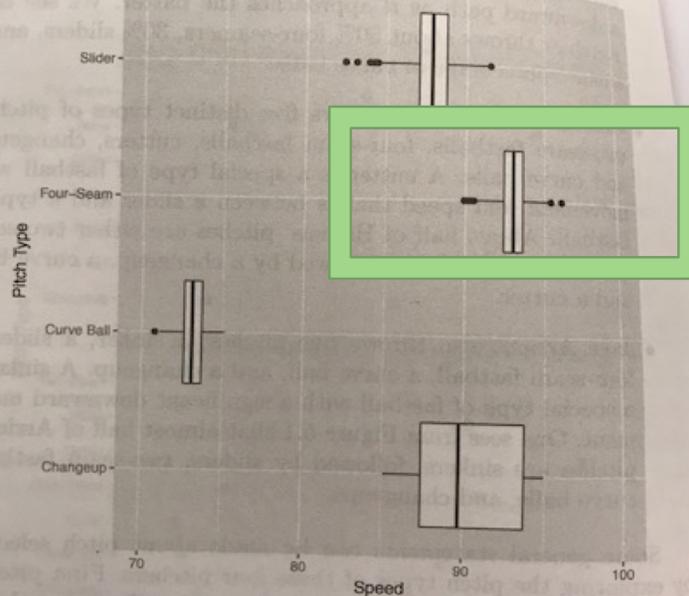
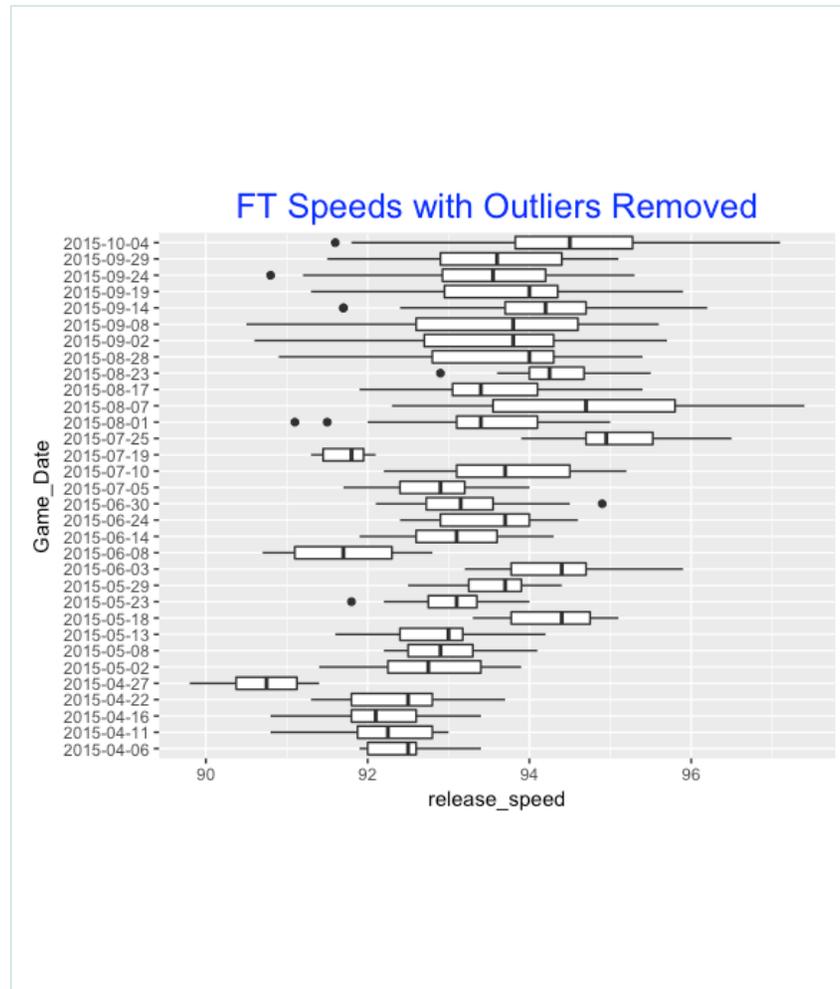


Figure 5.2 Boxplots of speeds of pitches of different types thrown by Clayton Kershaw in the 2016 season.

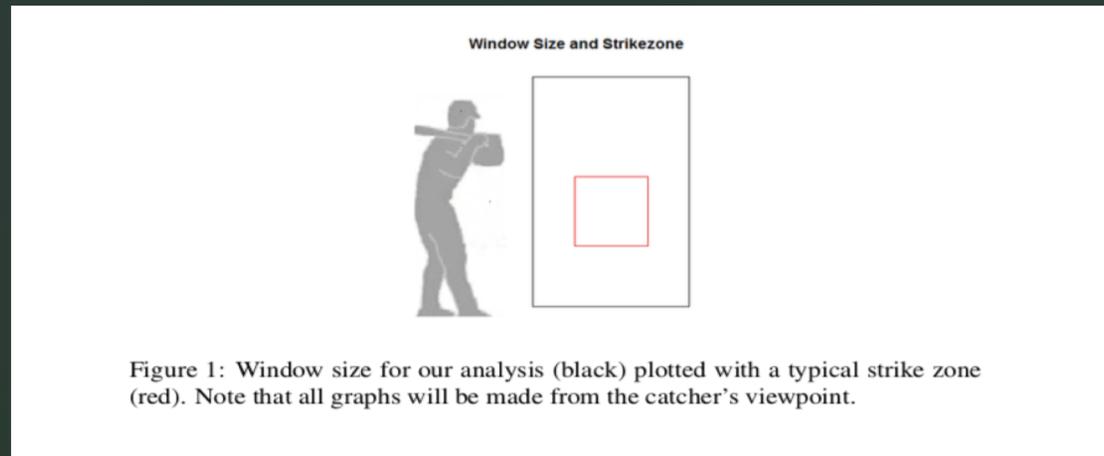
- **Helpful for comparing types of pitches & when to use them**
 - Fastballs over 90 mph are far less likely to be hit for home runs than swung on and missed
- **Helpful for understanding batter's hit/miss statistics**
 - **Example:** What's depicted here?
 - Kershaw's four seam fast ball is the fastest and the curve ball is the slowest
 - If a batter is expecting Kershaw to throw a four seamer and actually he throws a curve ball instead, visualized Pitch f/x can show us how the batter was fooled based upon speed

PITCH f/x Classify Pitch Type: Speed

- **Helpful for comparing types of pitches over the season**
 - Why is this important? Because it can inform pitching staff what it working and what needs improvement (or who needs to be traded)
- **Example:** What's depicted here?
 - Progression over the season for Hamels
 - Half of the total variation in Hamels' FT speeds was due to variation between games. There is no "hot" and "cold" — instead a general increase in Hamels' pitch speeds during the 2015 season.



▶ PITCH f/x Classify Pitch Type: Location



- **Strike Zone:** the zone that cover the home plate and is btw the midpoint of the batter's torso and the hollow beneath his knee
- Where the pitcher throws within the Strike Zone determines whether it is a hit, strike or ball
 - Those thrown to the middle are easy to hit and those thrown to the edges
 - This framework is important for visualizations that tell us why batters swing and miss on particular pitch types

PITCH f/x Classify Pitch Type: Location of Pitches with Scatterplots



- Can show you the location of a pitch over the season or game
 - Catchers can recommend best pitch

Example: What's depicted here?

- Curve balls fall low & outside of the zone, so they are harder to hit
- Fastball thrown to the middle are easy to hit

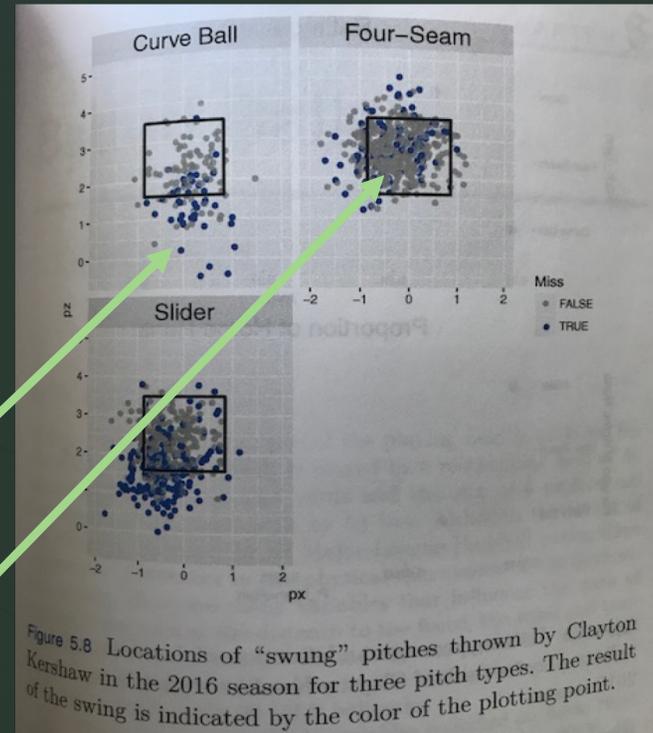


Figure 5.8 Locations of “swung” pitches thrown by Clayton Kershaw in the 2016 season for three pitch types. The result of the swing is indicated by the color of the plotting point.



Pitchers with unsuccessful appearances can (with their manager and pitching coach) look what went wrong

Media & Pitch Location Visuals

- Can show to the viewer the location of each pitch for a batter
- **Policy Connection-** By understanding where pitches are hitting in the zone, and comparing it to umpire decisions, MLB can use hard data to make helpful decisions in hypothetical rule changes

Example- If umpires called more high strikes, how many fewer foul balls and hits would there be in the course of a season? And to what extent could that reduction in batted balls further reduce the average time of a game?



▶ PITCH f/x Classify Pitch Type: Location Broken down by RH vs LH hitters

Pitchers will approach right-handed hitters (RH) and left-handed hitters (LH) differently and we can look at how their pitch type changes in these situations

Risk of over-plotting, points stacking on one another & attention being drawn to outliers

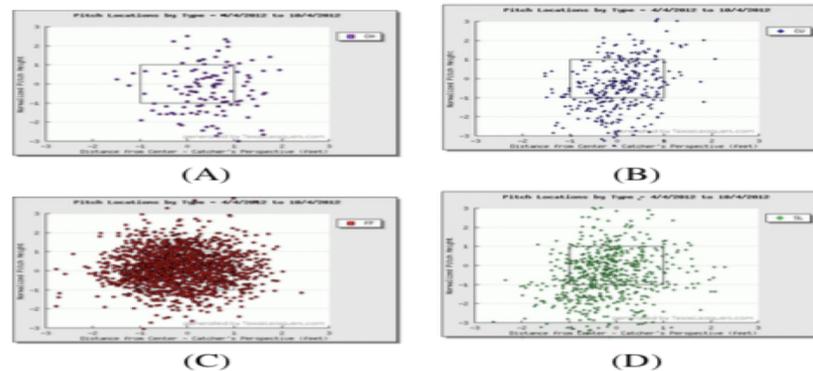
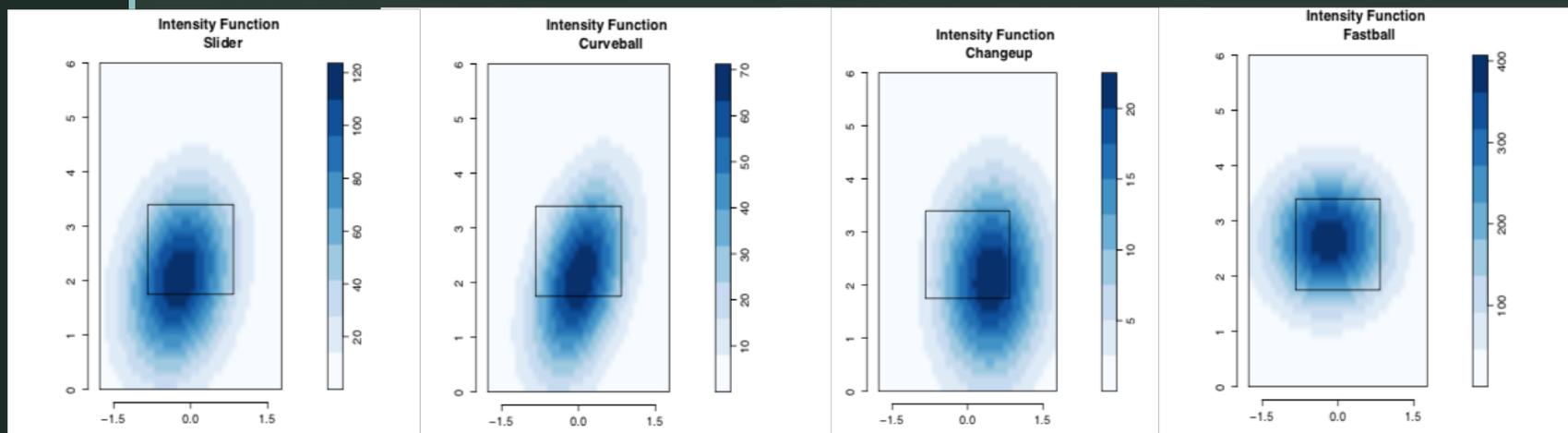


Figure 3: Scatter plots of Clayton Kershaw's pitches to RH hitters broken down by his four different pitch types: Changeup (A), Curveball (B), Fastball (C), and Slider (D)

Potential Solution: we can also estimate the point density across the plot area and indicate regions of different point densities with contour lines and colour (Carr et al., 1987).

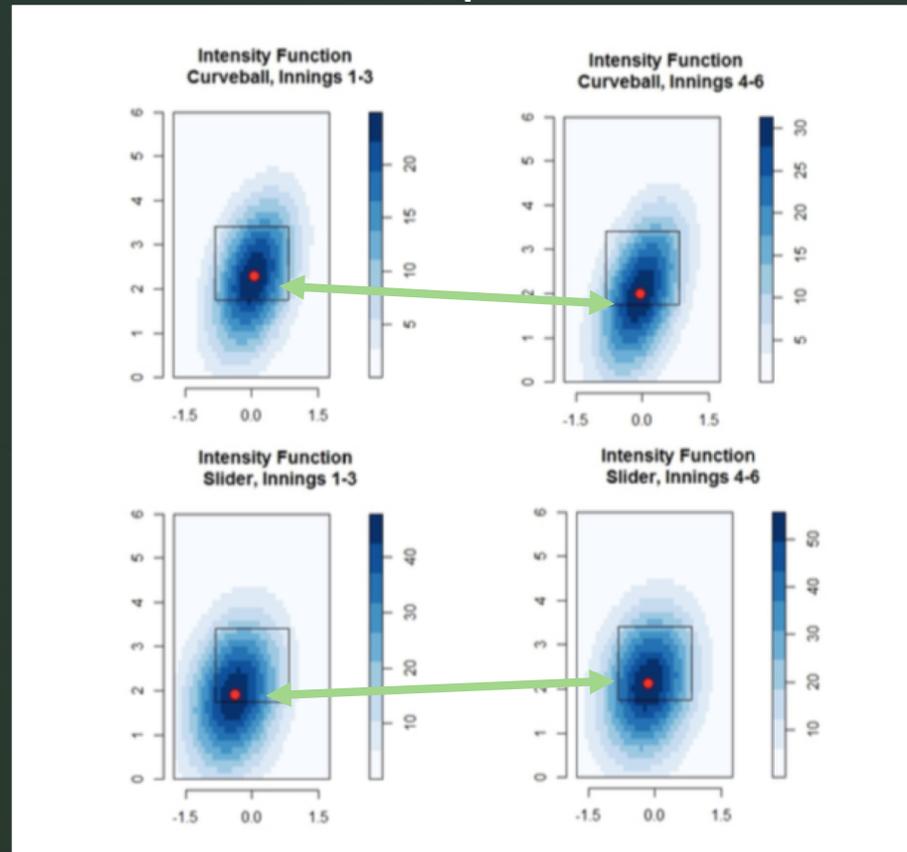
► PITCH f/x Classify Pitch Type: Location of Pitches with Heat Maps with RH



- Used to show how frequently a pitcher throws in each part of the strike zone
- The sequential color scale represents the magnitude of the intensity function
 - Some graphs will also use qualitative colour scales to make patterns more apparent
- Useful for evaluating trends in location with large sample sizes
 - *this visualization makes it harder to determine the exact data values shown*

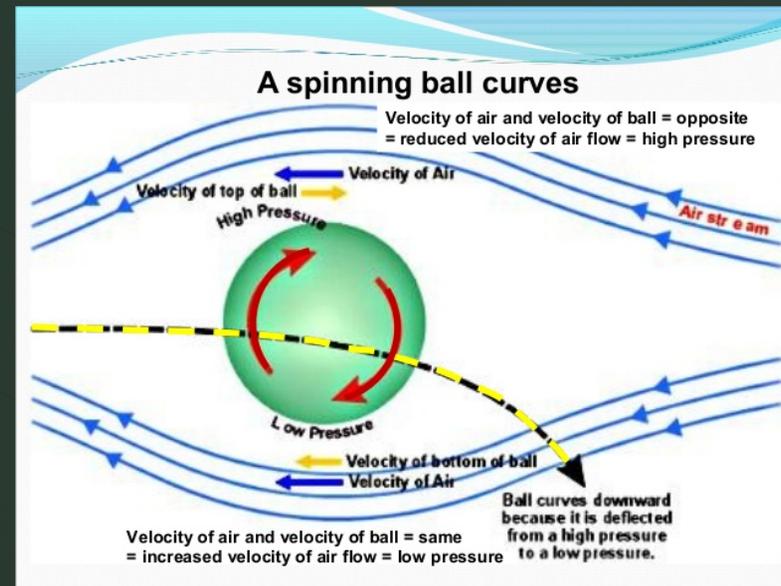
PITCH f/x Classify Pitch Type: Location of Pitches with Heat Maps with RH

- We can track how a pitcher's location is changing over the course of the game
 - Understand how a pitcher approaches his game
- **Example:** What's depicted here?
 - For the Slider - middle innings this pitch is higher in the strike zone and easier for a RH batter to hit.
 - Reason why Clayton Kershaw gives up more runs in the middle inning



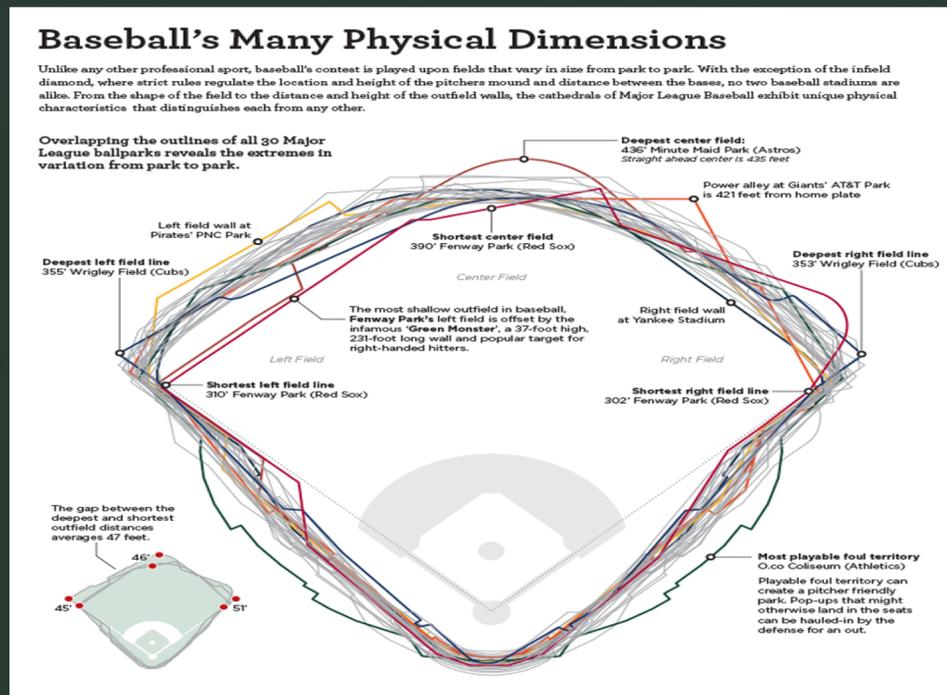
Fast (2010) on the Reliability of the Data & The Impact on Visuals

- Initial Experimentation
 - initial point of the pitch trajectory changed over time of the system use
- Evolving Algorithm
 - As the season goes on, a pitcher's pitch type information often becomes more accurate



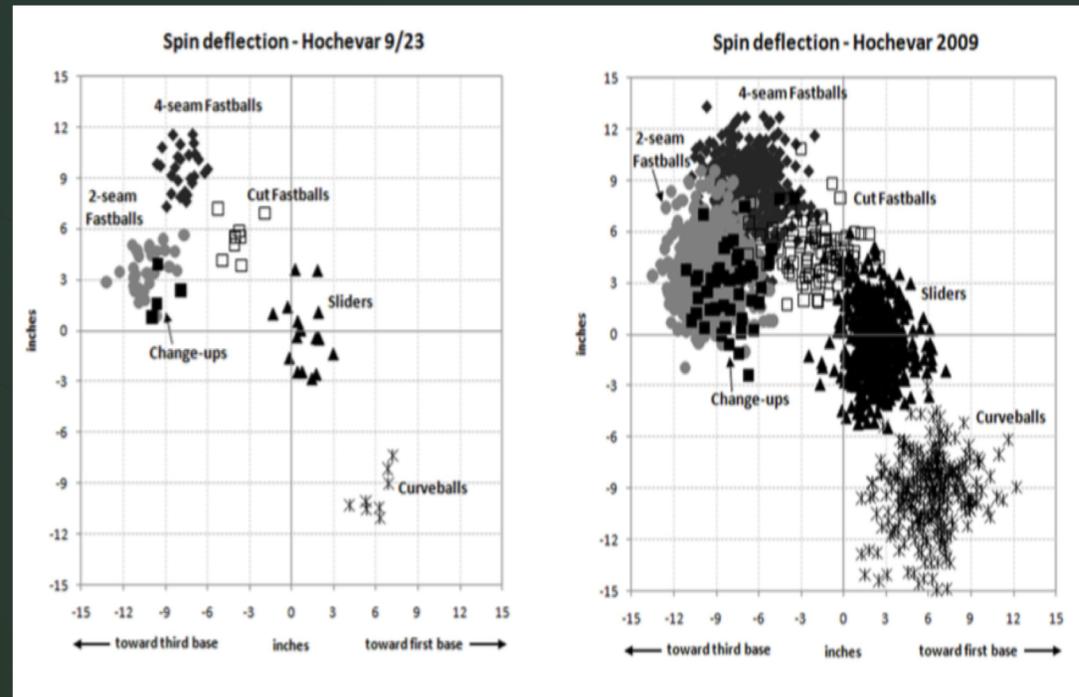
Fast (2010) on the Reliability of the Data & The Impact on Visuals

- External Factors
 - Park/ Camera System Bias
 - Air Pressure Differences
 - Air density affects how a pitched baseball moves



Fast (2010) on the Reliability of the Data & The Impact on Visuals

Visual complexity may hinder a person's ability to get a quick overview of the information or make it difficult to distinguish small differences in values. In general, we want to strike a balance between the two extremes and make our figures both memorable and clear (Bateman et al, 2010).



“inconsistencies between PITCH f/x system at different parks and sometimes at the same park at different points in the season, [trends] can be difficult to distinguish...”
For this reason, Michael Fast prefers to identify pitches on a game level and not a season level.

▶ To Find Your Own Baseball Data and/or Visuals

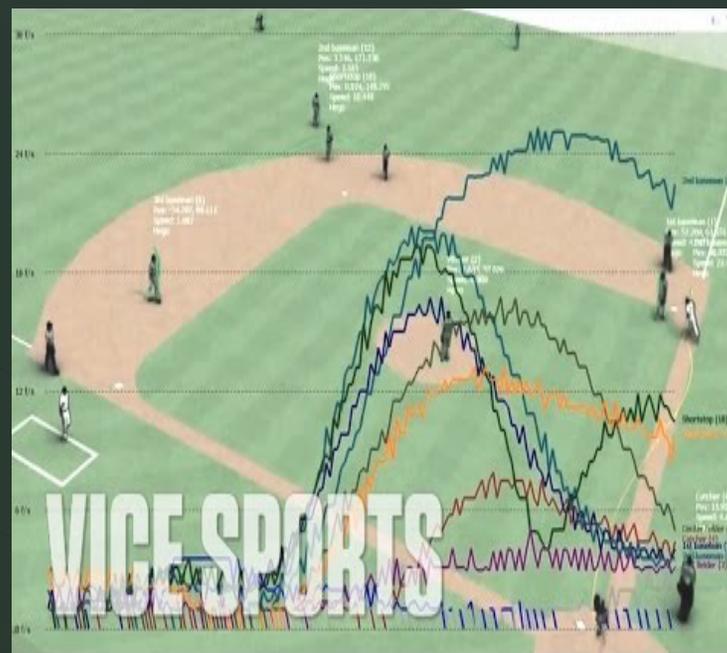
- **Joe Lefkowitz' tool**
 - allows you to download the data of a specific pitcher.
- **Brooks Baseball**
 - allows you to download excel data for a pitcher's numbers for a specific start
- **Fangraphs**
 - carries average data - horizontal and vertical movements as well as pitch velocity, for every pitch of a pitcher.
 - carries pitchf/x charts for each game of a pitcher as well as graphs for a full season's worth of pitches.
 - the only place where locational heat maps can be found as far as I know, which can be useful.
- **Texasleaguers**
 - contains the basic average numbers for pitches and relevant pitch results such as swing rates.
 - includes some relevant graphs which are quite nice.
- **Baseball Reference**
 - **Tracks player stats much more complexly**
- **Brooks Baseball**
 - notable for being the only one of these sites to update after each batter in a game.

Going Beyond PITCH f/x: The Evolution of Statcast

- In 2017, PITCH f/x was phased out for Statcast
 - Tracks the ball (and players) using a combination of radar AND cameras.
- Strong critiques similar to when PITCH f/x first introduced
- Has contributed to further leaps in our understanding of pitch analysis
- **What is it doing better?**
 - By contrast, Statcast measures right out of the pitcher's hand.
 - Which means, readings will nearly always be faster

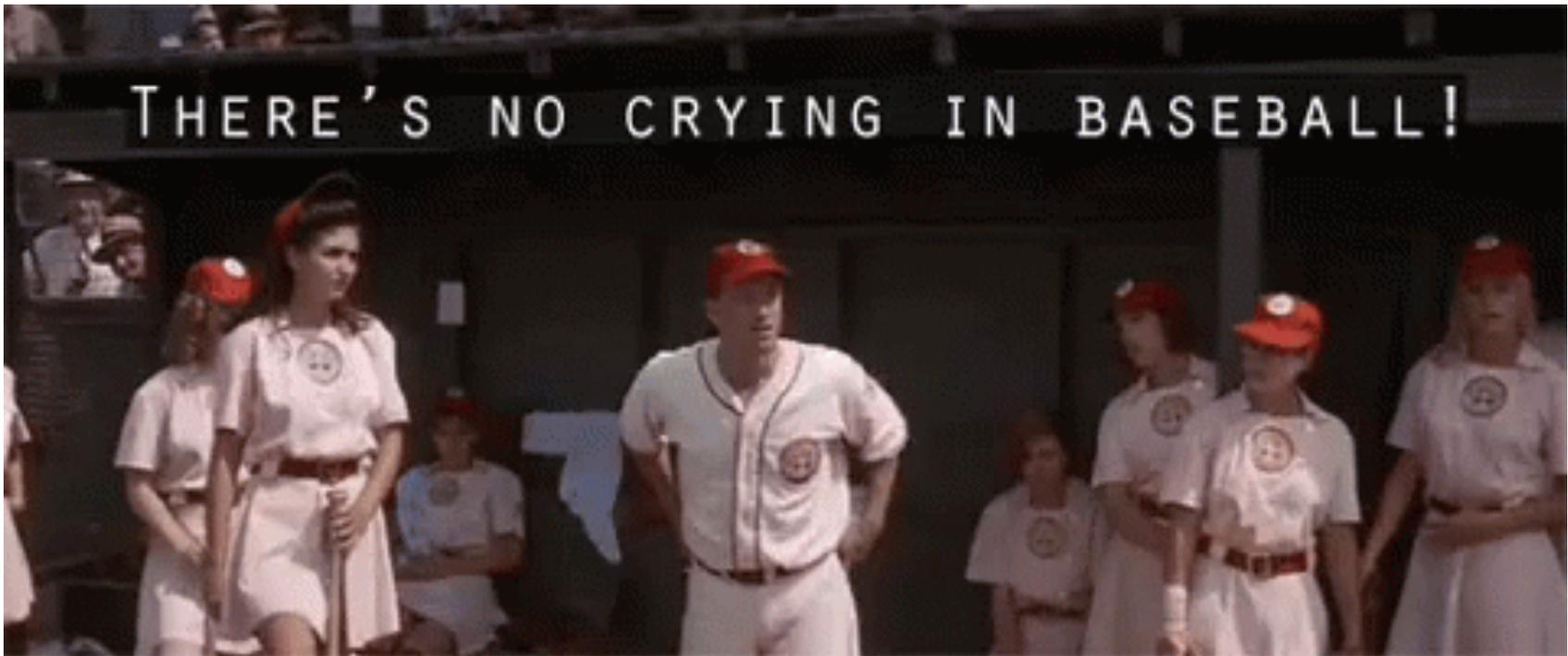
Controversial beginnings

<https://www.sporttechie.com/major-league-baseball-sued-pitchfx-system-statcast-sportvision-sportsmedia-technology-corporation/>



<https://www.youtube.com/watch?v=9rOKGKhQe8U>

THERE'S NO CRYING IN BASEBALL!



References

- Albert, J. (2018). *Visualizing Baseball*. New York: Chapman and Hall/CRC, <https://doi.org/10.1201/9781315149530>
- Albert, J., & Bennett, J. (2007). *Curve ball: Baseball, statistics, and the role of chance in the game*. Springer Science & Business Media.
- Bateman, S., R. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. 2010. "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts." *ACM Conference on Human Factors in Computing Systems*, 2573–82
- Carr, D. B., R. J. Littlefield, W. L. Nicholson, and J. S. Littlefield. 1987. "Scatterplot Matrix Techniques for Large N." *J. Am. Stat. Assoc.* 82: 424–36.
- Fast, M. (2010, April 18). A PITCHf/x primer. Retrieved January 28, 2019, from <https://fastballs.wordpress.com/2010/04/18/a-pitchfx-primer/>
- Garik16. (2011, March 31). A Pitch F/x Primer: Some tips on Run Values, Pitch Classifications, and Heat Maps. Retrieved January 29, 2019, from <https://www.beyondtheboxscore.com/2011/3/31/2068855/pitch-fx-primer>
- Kalkman, S. (2009, April 17). Understanding Pitch f/x Graphs: Location vs. Movement. Retrieved January 28, 2019, from <https://www.beyondtheboxscore.com/2009/4/17/841366/understanding-pitch-f-x-graphs>
- Long, J. (2014, July 22). Why is PITCHf/x important. Retrieved March 24, 2019, from <https://www.beyondtheboxscore.com/2014/7/22/5919581/why-pitchfx-is-important>
- Mills, B. M., & Sievert, C. (2017). Using publicly available baseball data to measure and evaluate pitching performance. In *Handbook of Statistical Methods and Analyses in Sports* (pp. 55-82). Chapman and Hall/CRC
- Seeker. (2015, June 28). The Physics Of Baseball Pitches. Retrieved January 29, 2019, from <https://www.youtube.com/watch?v=H09TIJ4a4fg>
- Sports, V. (2015, September 25). Future of the Game: Baseball's Latest Statistical Revolution. Retrieved January 29, 2019, from <https://www.youtube.com/watch?v=9rOKGKhQe8U>
- Wilcox, A., & Mannshardt, E. (2013). *Baseball scouting reports via a marked point process for pitch types*. North Carolina State University. Dept. of Statistics.