Luncheon Keynote

How Data Science Changed Major League Baseball: What the Gaming Industry Can Learn from Batter Up!

Ari Kaplan

Sports Analytics

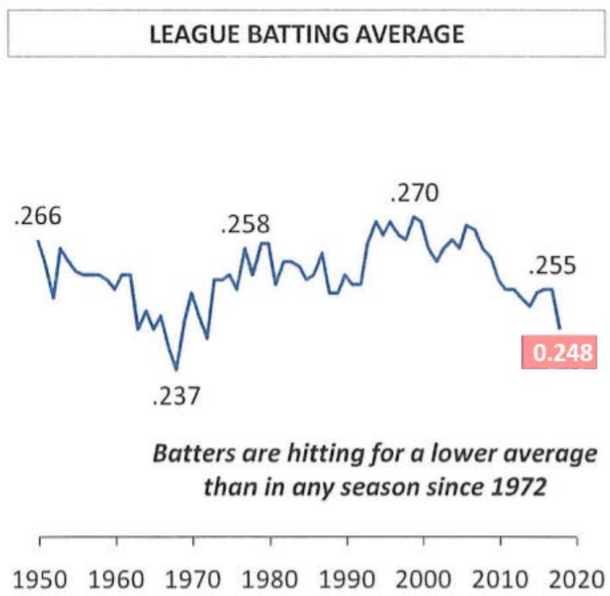
Monday, January 28
11:45 am – 1:15 pm

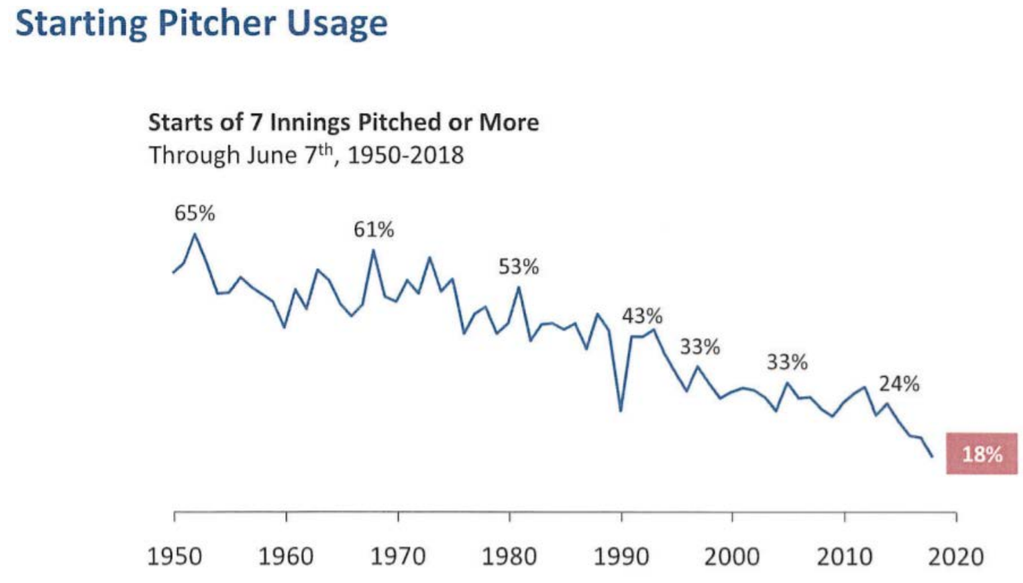
RavingNEXT

Contact Ari Kaplan:
kaplanari@hotmail.com or connect with him on LinkedIn

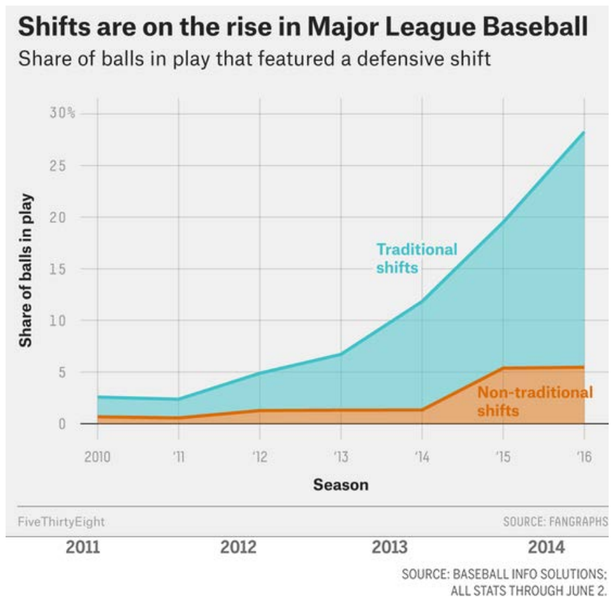
Analytics changing how the game is played

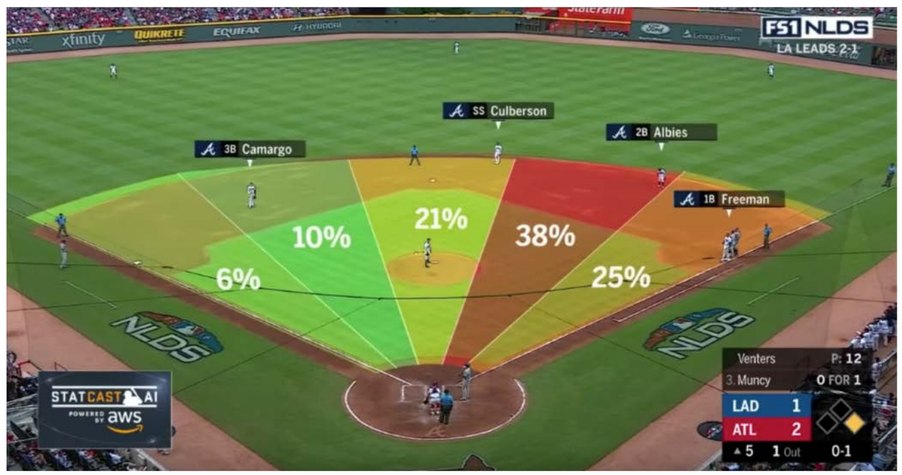
Highest SO% in history, lowest BA since 1972





Most shifts in history





Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

“THE GAME” and Gaming Industry

MGM Resorts will also gain access to MLB’s treasured statistical data, albeit on a non-exclusive basis. However, some enhanced stats will be provided to MGM on an exclusive basis.

M.L.B., Once Averse to Gambling, Strikes a Deal With MGM Resorts

On Tuesday, James Murren, chief executive of MGM Resorts, and Rob Manfred, M.L.B.’s commissioner, announced MGM becoming baseball’s official gambling industry partner. Seth Wenig/Associated Press

DraftKings Gives NFL Fans First Taste Of In-Stadium Sports Betting In U.S.

FANDUEL

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

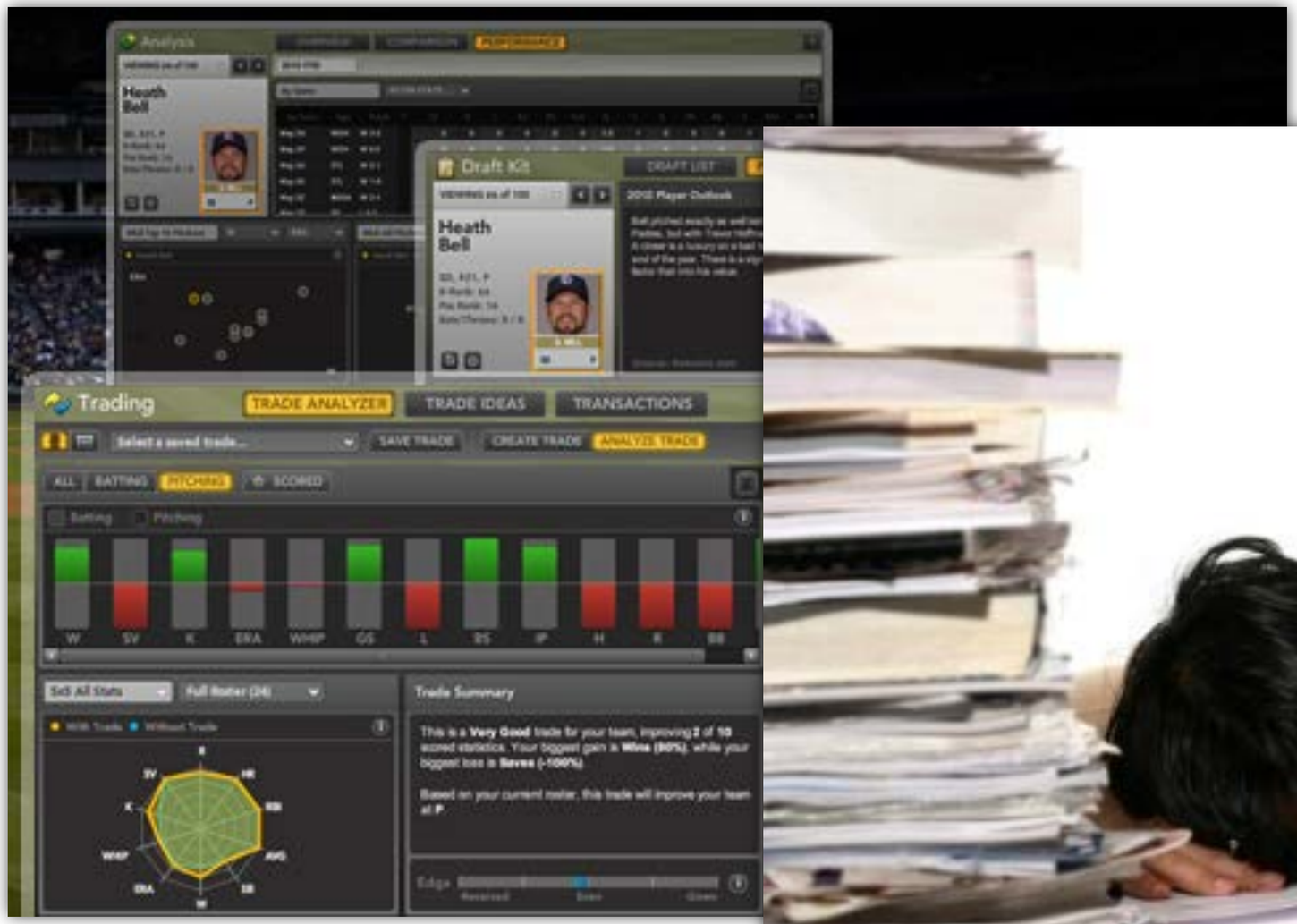
Win!

“What’s more important than the will to win is the will to PREPARE to win”

– Coach Wooden

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Prepare to win!



Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Seek open-ended questions …

Why is a batter struggling?
Why aren't we attracting gamers as much as we thought?
Is an injury affecting a pitcher's mechanics?
What would happen if we lowered our F&B prices?
What should our next steps be?

… then give actionable answers.

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Apply analytical models again and again to get answers

Above the Field

Player forecasting
Economics of contracts
Roster modeling
Trade-deadline analysis
Drafts

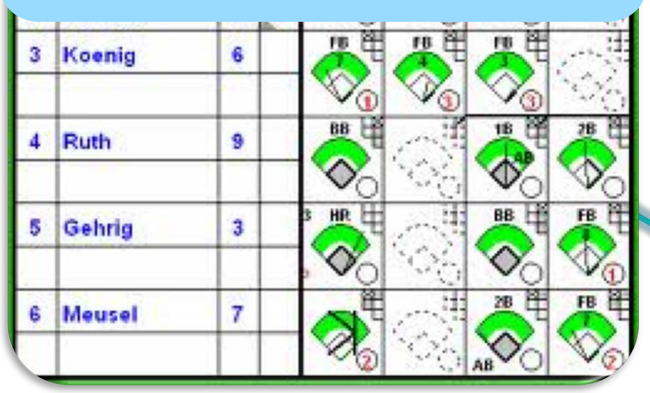
On the Field

Game preparation: advance scouting
Pro-scouting
Amateur scouting
International scouting
Player development
Injury prediction and management

RavingNEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Structured data sources (externally collected)

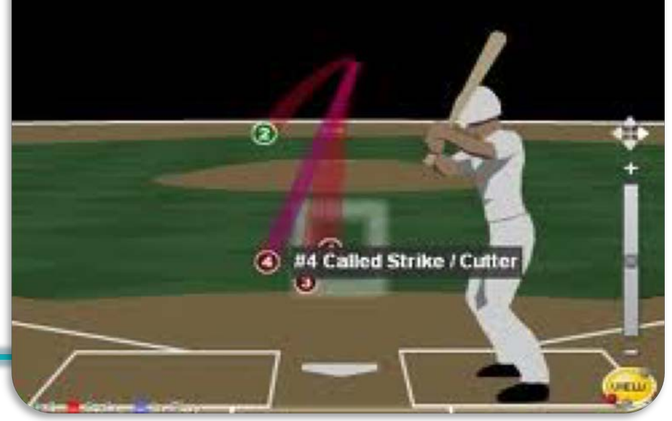
Play-by-play (MLB Advanced Media)



Statistics (MLB, STATS, AriBall, Inside Edge)

Defense (BIS, Statcast)

Pitch & hit mechanics (SportVision, Trackman)



Contracts & Financials (eBIS, Stadium operations)

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Property of Ari Kaplan. Copyright 2019

Multi-structural data sources (internally collected)

Organizational Video

MiLB Coach Reports

Pro-Scouting Reports

Amateur Scouting Reports

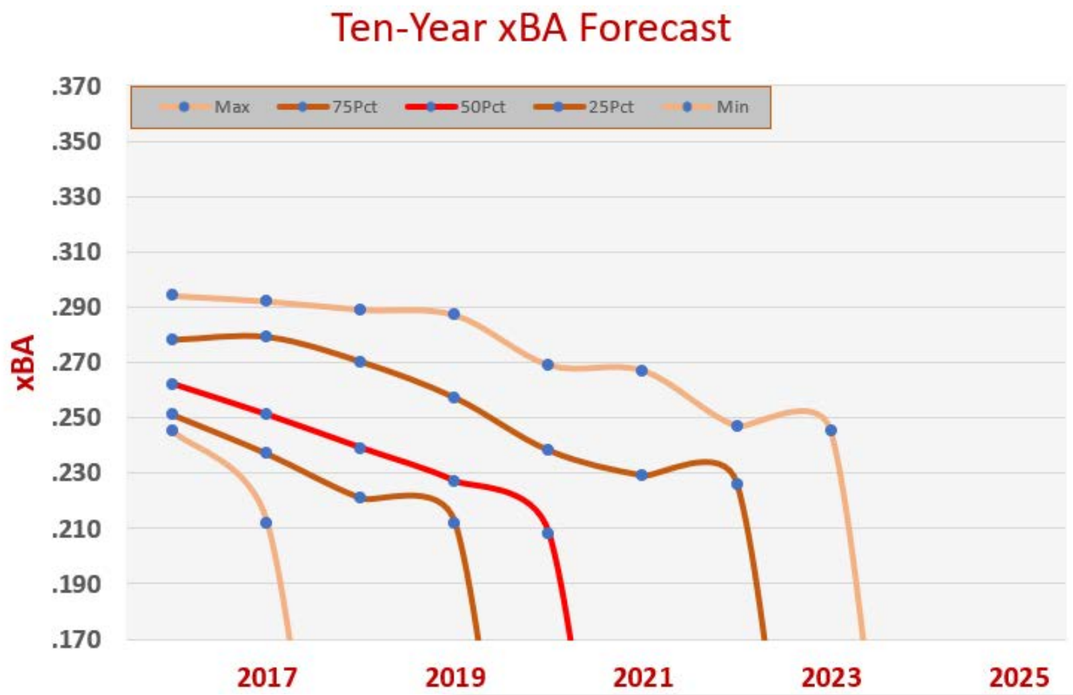
Injury & Medical

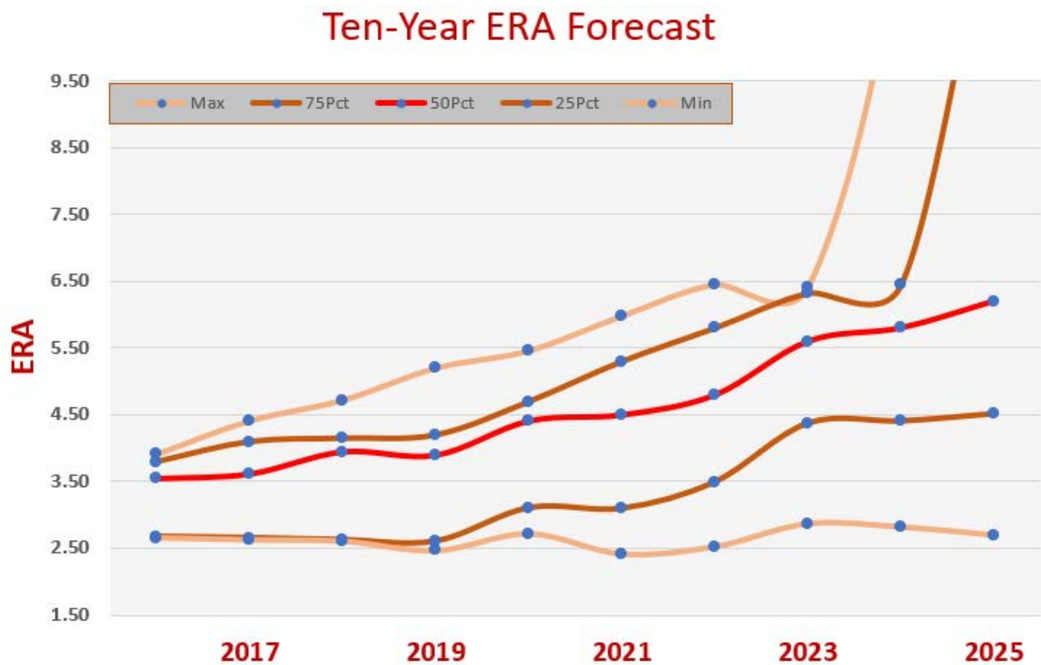
Advance Scouting Reports

International Scouting Reports

RavingNEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Understand and manage risks





MGR [redacted]

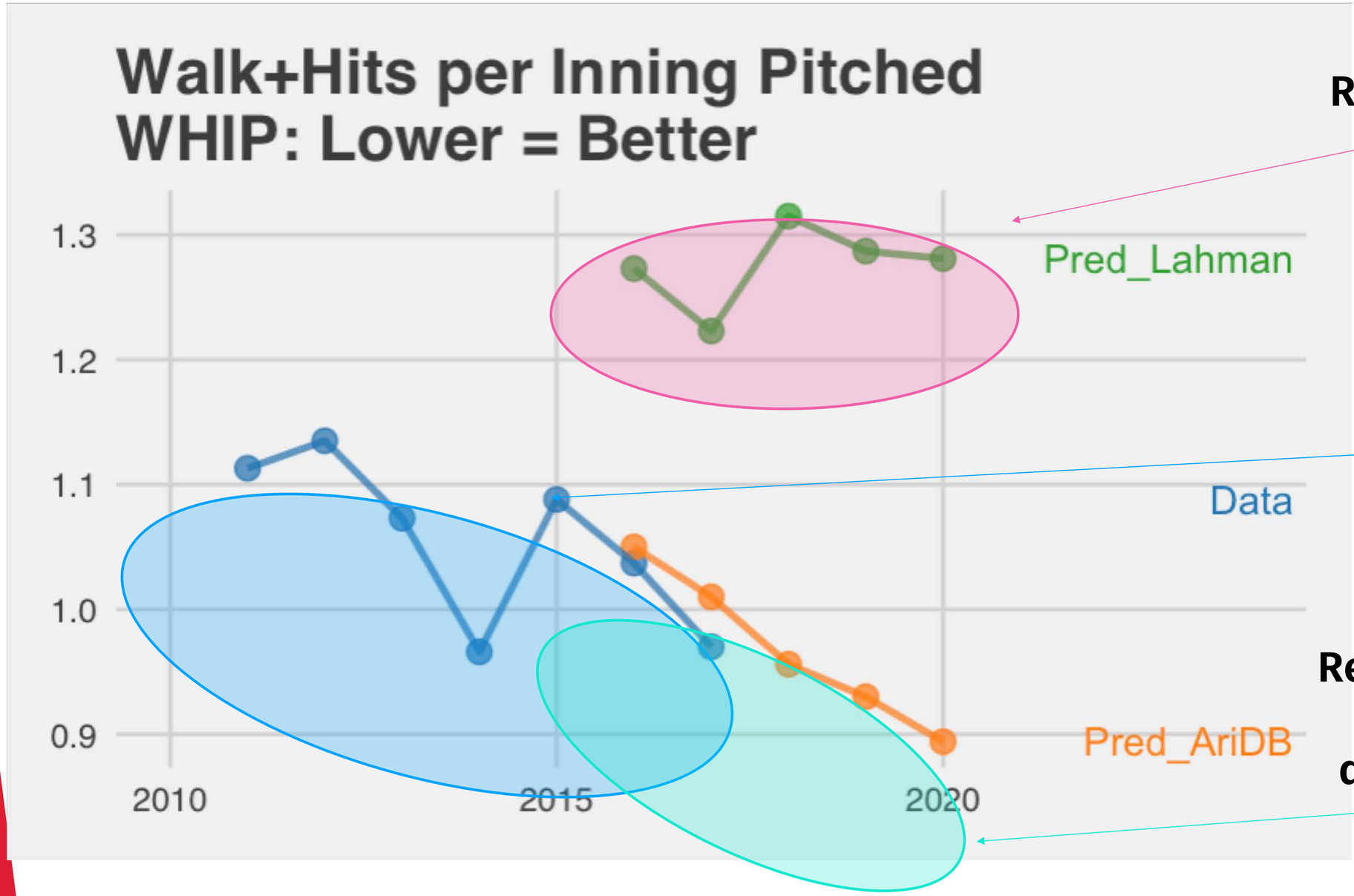
Pos	Player	Age	AB	OBP	SLG	Bat RC
1-SS	[redacted]	21	575	0.359	0.428	85
2-CF	[redacted]	33	475	0.339	0.428	67
3-1B	[redacted]	33	475	0.354	0.473	76
4-3B	[redacted]	33	475	0.342	0.498	78
5-LF	[redacted]	35	450	0.317	0.474	67
6-RF	[redacted]	25	425	0.307	0.471	61
7-C	[redacted]	28	425	0.373	0.486	72
8-2B	[redacted]	25	425	0.327	0.372	51
9-P	pitchers		300	0.170	0.159	1
OF	[redacted]	34	400	0.367	0.431	61
UT	[redacted]	29	325	0.336	0.448	48
C	[redacted]	31	200	0.287	0.343	20
IF	[redacted]	25	175	0.294	0.291	15
OF	[redacted]	29	125	0.368	0.344	16
	others		250	0.285	0.325	23
	totals		5500			740

	Pitcher	Age	IP	ERA	RA
SP	[redacted]	30	221	3.67	**98**
SP	[redacted]	34	214	3.83	**99**
SP	[redacted]	28	201	4.12	**100**
SP	[redacted]	32	133	4.80	**77**
SP	[redacted]	28	133	4.06	**65**
S/R	[redacted]	24	114	5.03	**69**
CL	[redacted]	28	83	3.14	**31**
RP	[redacted]	28	68	3.84	**31**
RP	[redacted]	34	62	3.48	**26**
RP	[redacted]	32	50	4.28	**26**
RP	[redacted]	27	45	5.00	**27**
RP	[redacted]	26	34	5.03	**21**
RP	[redacted]	25	22	5.71	**15**
	others		60	5.63	**41**
	totals		1440		**726**

Forecast WPctg:	0.51
Forecast Wins:	83
Forecast Losses:	79

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Statistical Analysis: predicting from Lahman & Statcast



Results from models based on Lahman data only

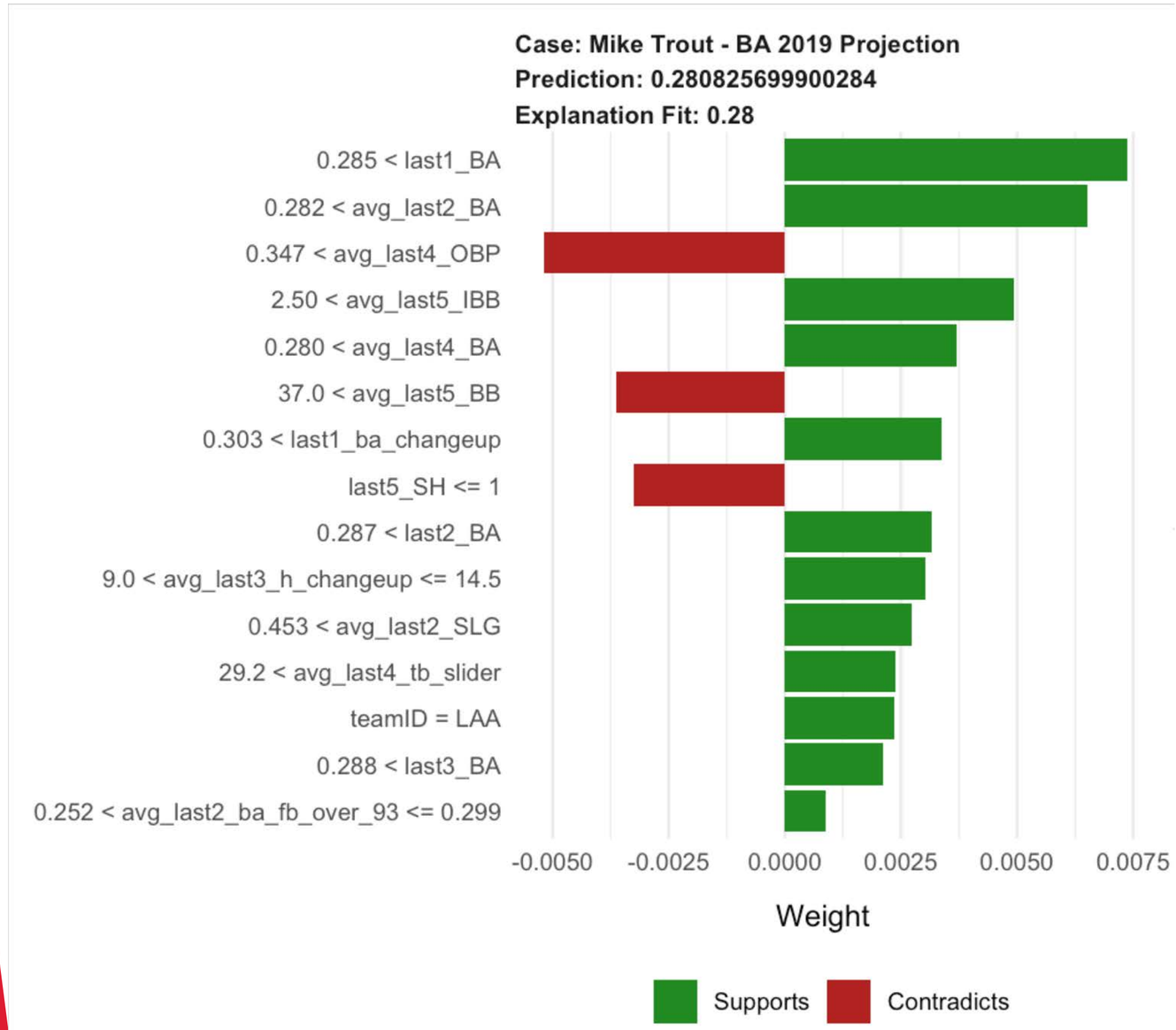
Historical player performance data

Results from models based on final dataset (Lahman + Trackman)

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Statistical Analysis: predicting from Lahman & Statcast: LIME

LIME – Local Interpretable Model-agnostic Explanations




```
# Install 'lime' from CRAN
install.packages('lime')
```

- Approximate reasoning of complex ML models (ensembles).
- Most important attributes and their contributions to the predictions.
- Validated the models with his baseball domain knowledge.

Refocus workers from the mundane to the strategic

What happened? What is happening? What will happen?

Scoutable™ reports: based on full coverage of every pitch, every game. The reports are presented in the same formats that many scouts and organizations use today.

- **Habits: threw FB whenever there was a 3-ball count. Never threw consecutive pickoff moves.**
- **Strengths: changeup had a big fading action. Kept first-pitches down 48% of the time (25% was avg). Plus control of his FB.**
- **Summary: Threw FB 91-94 (34% of all pitches), cutter 88-91 (12%), sinker 91-94 (28%), curveball 75-78 (13%), changeup 85-87 (13%)**
- **Last game compared to before: threw sinkers 16% less often and cutters 14% more often**

Text Analysis of Scouting for predictive analysis

Scouting Report

Pos Seen	LHS	Ht/Wt	6' 3" / 200	DOB	0 /21/199	*Pres/Fut Role	1 - 2
Future Pos	LHS	Bats/Throws	Left / Left	Age	23	Interest (given Role grade)	Low
MLS	0 + 000	Next Level	2A	High Future	ML	Conviction	6
Agent							

Physical Description long legs, sloped shoulders

Scout Name	Entered	Team Name	Organization	Level
	0 /12/201	[redacted]	[redacted]	1A

Pitcher Evaluation

Date of Last Game Seen	05/07/2015
Games Seen	2.0

Innings Seen	14.0

Fastball			Radar			Miscellaneous (other than Out Pitches/Optimal Role, all grades should be present grades)	
Tools	Present Abilities	Future	Low	High	Comfort Zone		
Fastball ()	4	4	84	91	89	Arm Angle	High 3/4
Movement in the Zone	4	4				Arm Action	Plunge, Quick, Recoil, Wrapper
Fastball Command	5	5				Direction (stride)	Straight
FASTBALL VALUE	4	4				Direction (toe)	Straight
						Delivery	Plus
						Plane	Avg

Secondary Pitches			Radar		Command			
Tools	Present Abilities	Future	Low	High	Present	Future	Deception	Avg
Curve (Down, Roller)	4	4	68	74	5	5	Pitchability/Feel	Plus
Slider (Slurve, Tilt)	3	3	77	81	3	3	Overall Athleticism	Avg
Change (Circle)	5	6	76	79	4	5	Field	
Splitter ()							Holding Runners	Plus
Other ()							Release Times	1.12 (FB) | 1.16 (CB) | ()
							Slide Step Times	1.05 () | 1.09 () | ()
							Out Pitch RHH	Change
							Out Pitch LHH	Change
							Optimal ML Role	Emergency Extra

Overall Control	6	7

Overall Makeup:

Player Summary:
Pitched from far 3B side of the rubber in first start and was pitching from 1B side of the rubber a month later and much more effective from that angle. Balanced, consistent delivery with upright finish and some recoil. Delivery is not deceptive but efficient and repeatable. Everything from same arm speed. Doesn't throw balls. Jamie Moyer - Barry Zito type, with big curve ball as feature pitch. Decent athlete and plus poise on the mound. Mixes up pitches well and pitches to contact. Everything is around zone. Tougher on LHH, gets called 3rd strikes and fools them. Lots of pick-offs with average move and a quick, step-back pick. Incorporates a quick pitch too.

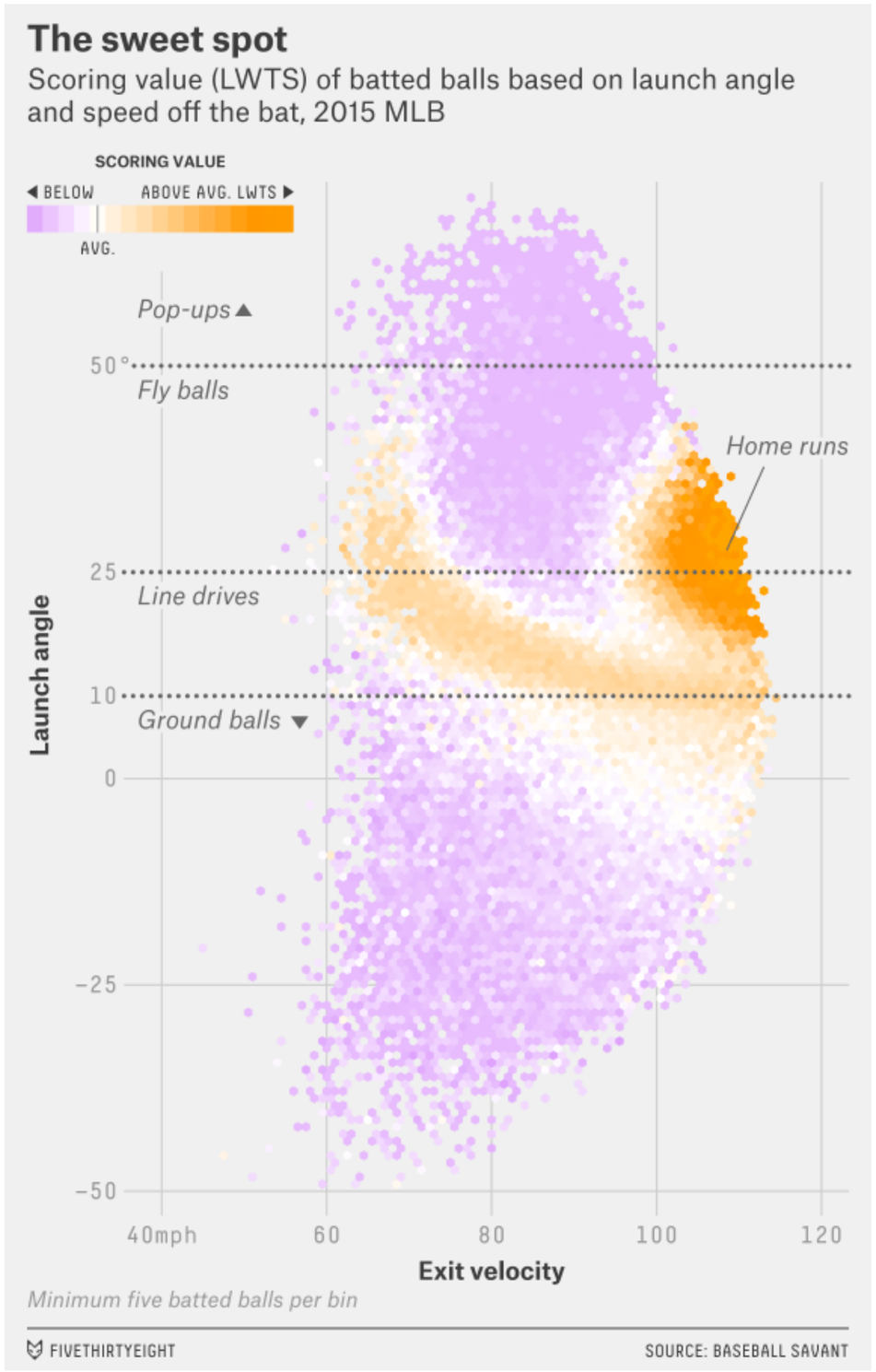
FB - tails, more horizontal action than vertical. In first start - 2 seam is 84-86 sits 85. 4 seam was 86-88, sitting 87 but very hittable. In second start FB was 88-91, sat 89-90. Doesn't maintain velo well as a starter. Hit a wall after 70 pitches (sat 83 in 6th inning of first start).
CB - big bending, swing and miss offering to RHH. Not much deception, relies on movement to be effective. Will start off LHH with it. Won't play up against hitters with avg to plus pitch recognition.
SL - Slurve has some tilt, tendency to leave up, worst offering. Junked it completely by 2nd outing.
CH - lots of weak contact against it. Will throw it first pitch to RHH and backs it up. Consistent to low and away location to both RHH and LHH. Will go low and in to LHH. Has drop and abv average movement with deception. Swing and miss offering to RHH. Sometimes has too much sink, also a little cut.

Type: Profiles as a finese pitcher with plus control and one abv average pitch (CH). Will do well against below avg minor league hitters that can't pick-up 3-4 pitch mix, and can be effective vs. better hitters with slow mix and avg deception, keeping hitters guessing and off-balance. At upper levels lack of velo and lack of abv avg off-speed command of hitting spots will be detrimental to advancement. Not especially tough on LHH. Best suited in long relief role out of pen with FB/CH combo.

One-Line Summary:
Upper minors finese longman, emerg extra at peak

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Goal of Statistics: separate value from luck



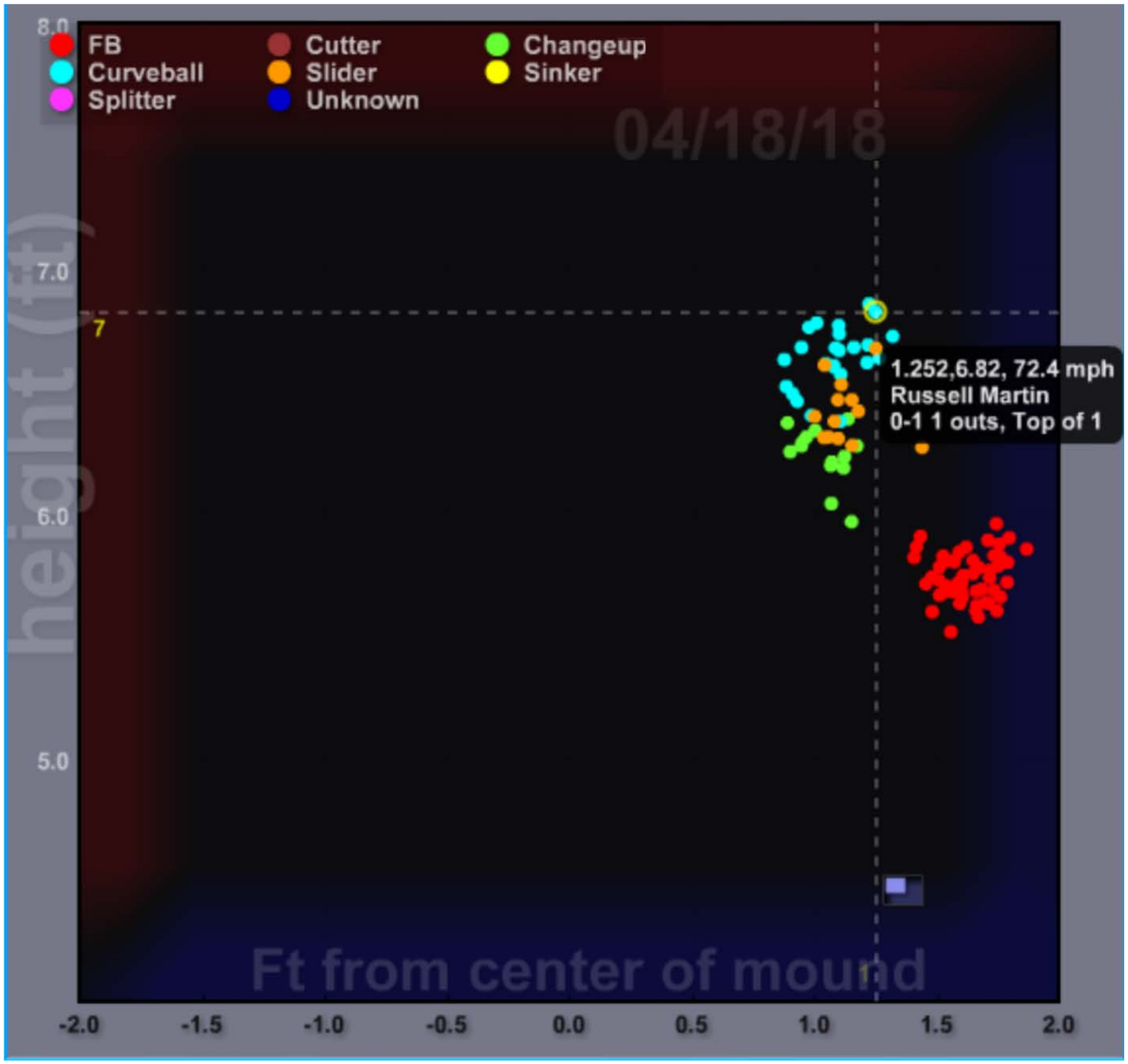
Hang Time	*#*	*Out%*
1.5 to 3.0	15,952	19.5%
3.0 to 4.0	11,749	58.3%
4.0 to 5.0	14,719	71.3%
5.0 to 6.0	16,717	84.7%
6.0 plus	6,281	93.3%

$$xFIP = \frac{13(xHR) + 3BB - 2K}{IP} + C$$

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Find actionable patterns in the data

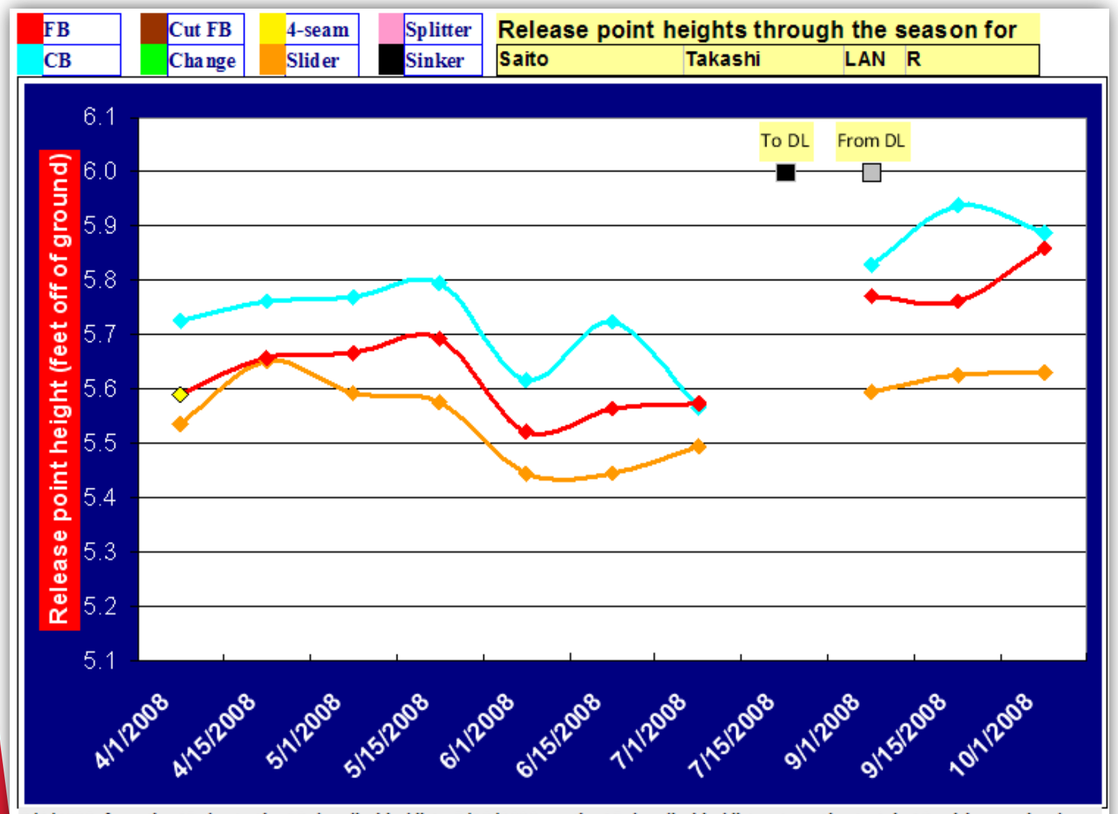
Fastball release points come down and over a foot



Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

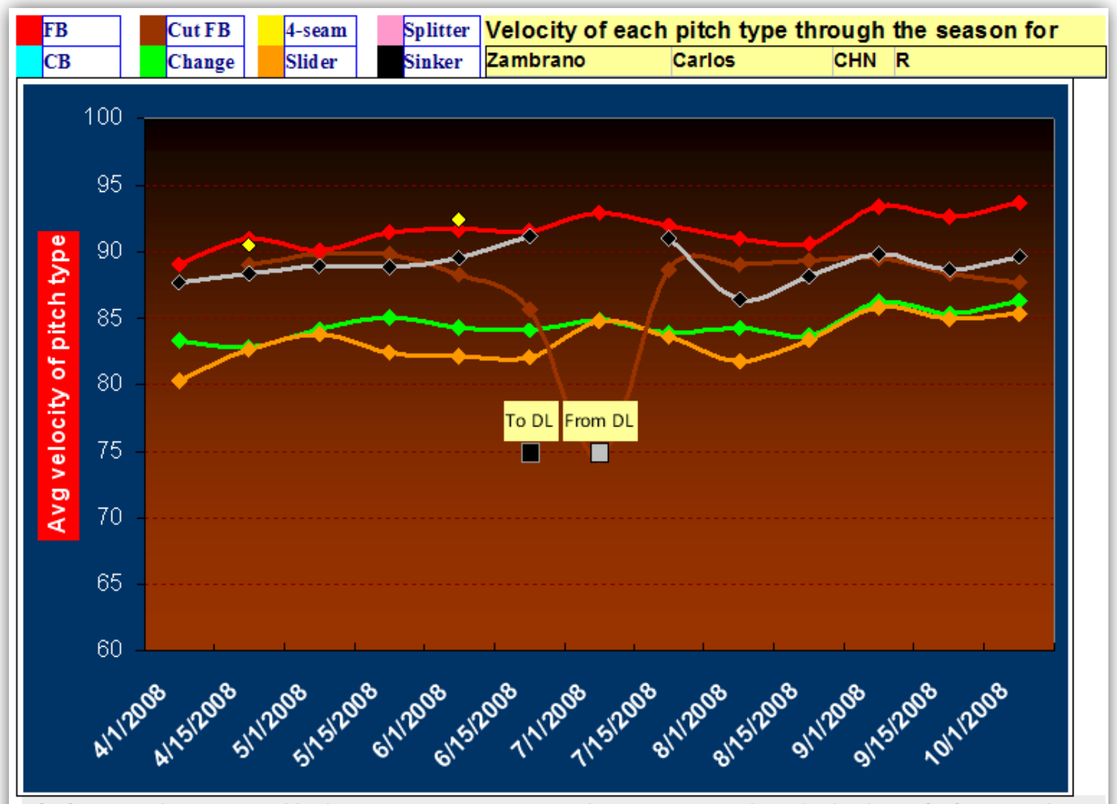
Find events preceding a business issue

Track consistency and deception of release points and velocity as the season progresses. See how injuries, trades, or assignments from the Minors affect mechanics.



9/4/2008: from the 15-day to the 60-day disabled list.. 7/18/2008: on the 15-day disabled list, retroactive to July 13, with a sprained ligament in his right elbow..

1/10/2009: Signed RHP Takashi Saito to a one-year contract with a club option for 2010.. 12/5/2006: Signed RHP Takashi Saito to a one-year contract.. 4/7/2006: Purchased the contract of RHP Takashi Saito from Triple-A Las Vegas. 3/23/2006: reassigned to Minor League camp.. 2/7/2006: Sent to to a Minor League.



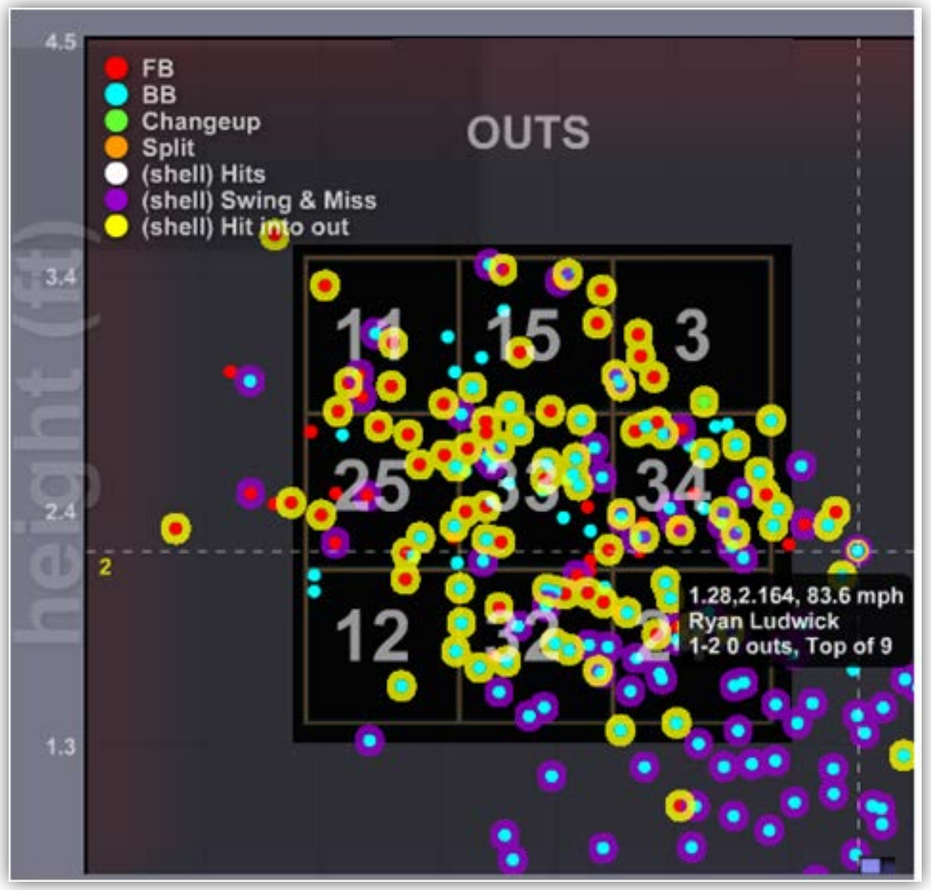
6/21/2008: on the 15-day disabled list, retroactive to June 19, with a minor strain in his right shoulder. 5/10/2002: on the 15-day disabled list with a strained right elbow..

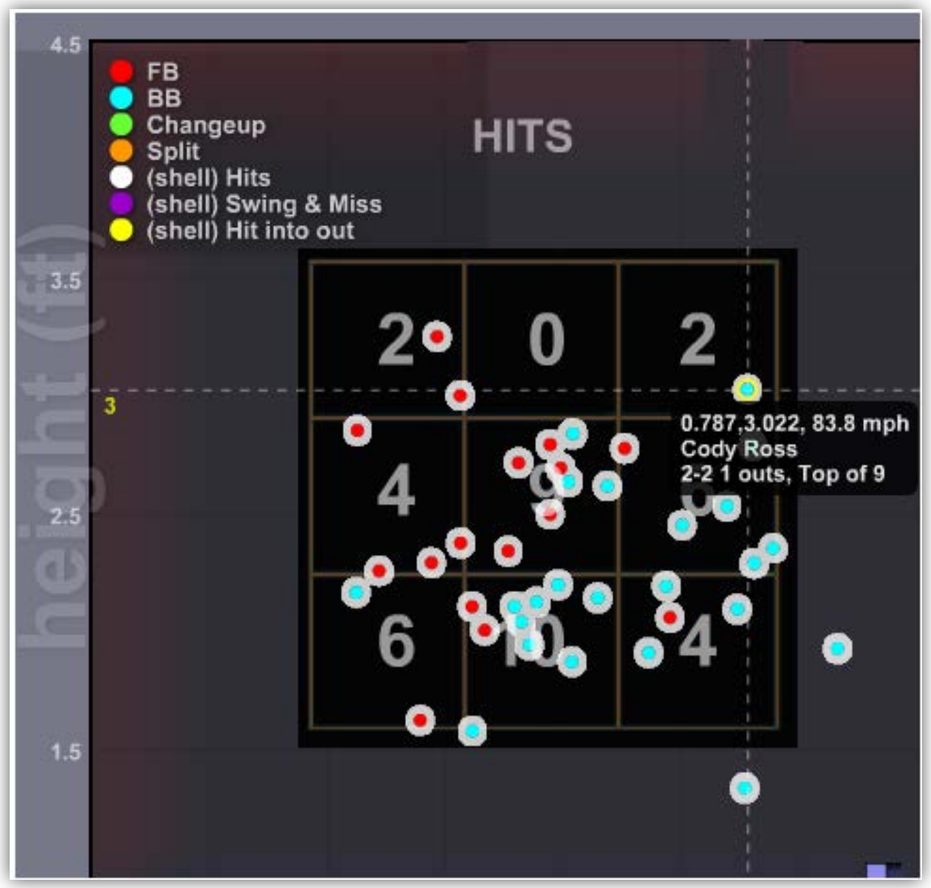
2/20/2007: Signed RHP Carlos Zambrano to a one-year contract, avoiding arbitration.. 2/4/2005: Agreed RHP Carlos Zambrano on one-year contract. 4/11/2002: The club recalled RHP Carlos Zambrano from Triple-A Iowa.. 8/20/2001: The Cubs purchased the contract of RHP Carlos Zambrano from Iowa (AAA). 3/26/2001: Sent to Minor League.

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Use intelligence to find opportunities

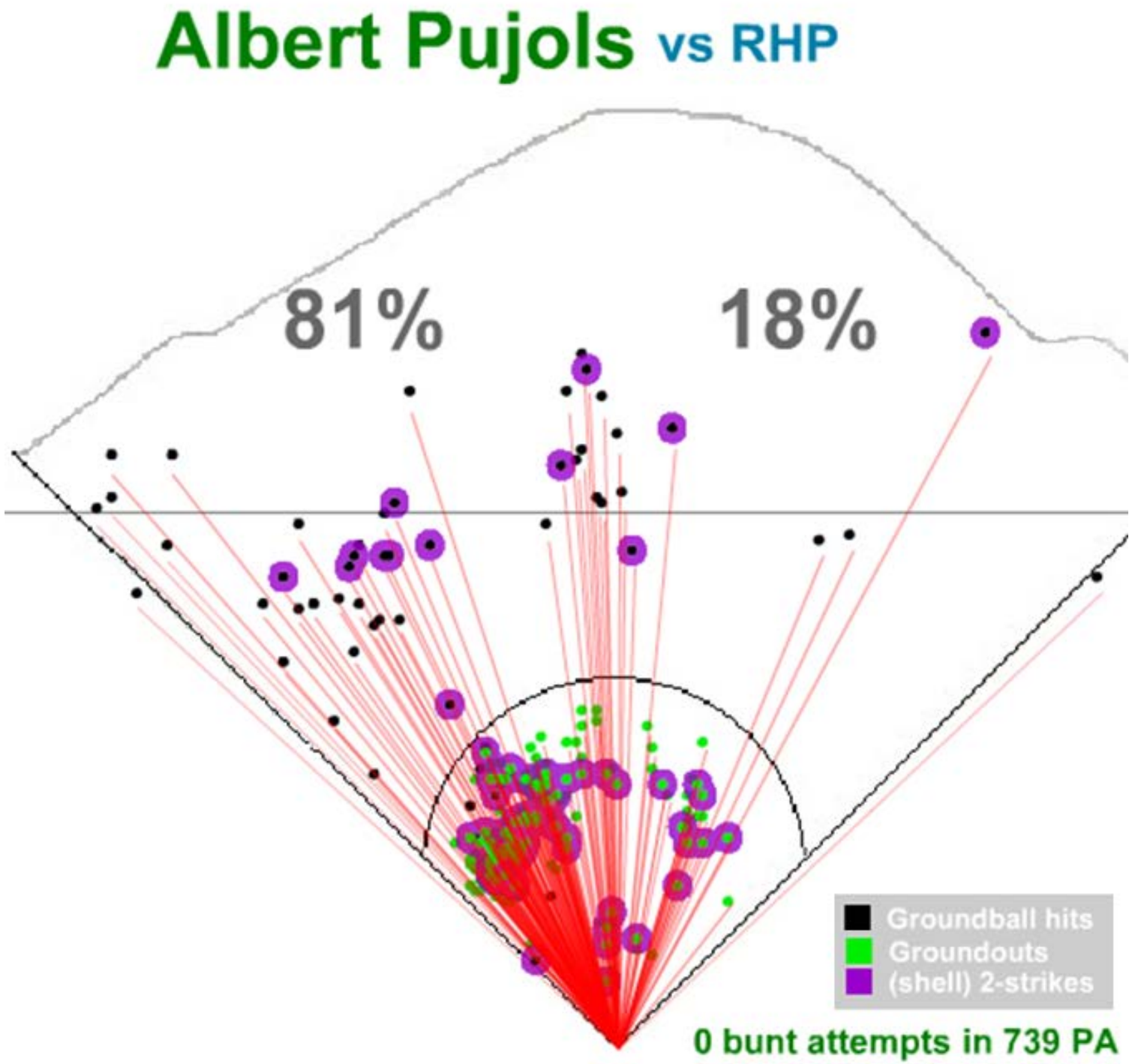
- Where did a pitcher generate outs? What pitch types and locations? Below shows FB up and slider low/away.
- Where did he allow hits? FB inner half, BB up in zone outer half.
- These reports can be for pitcher/batter matchups, across years, and much more





Use intelligence to better “defend” yourself

Be “positioned” to win

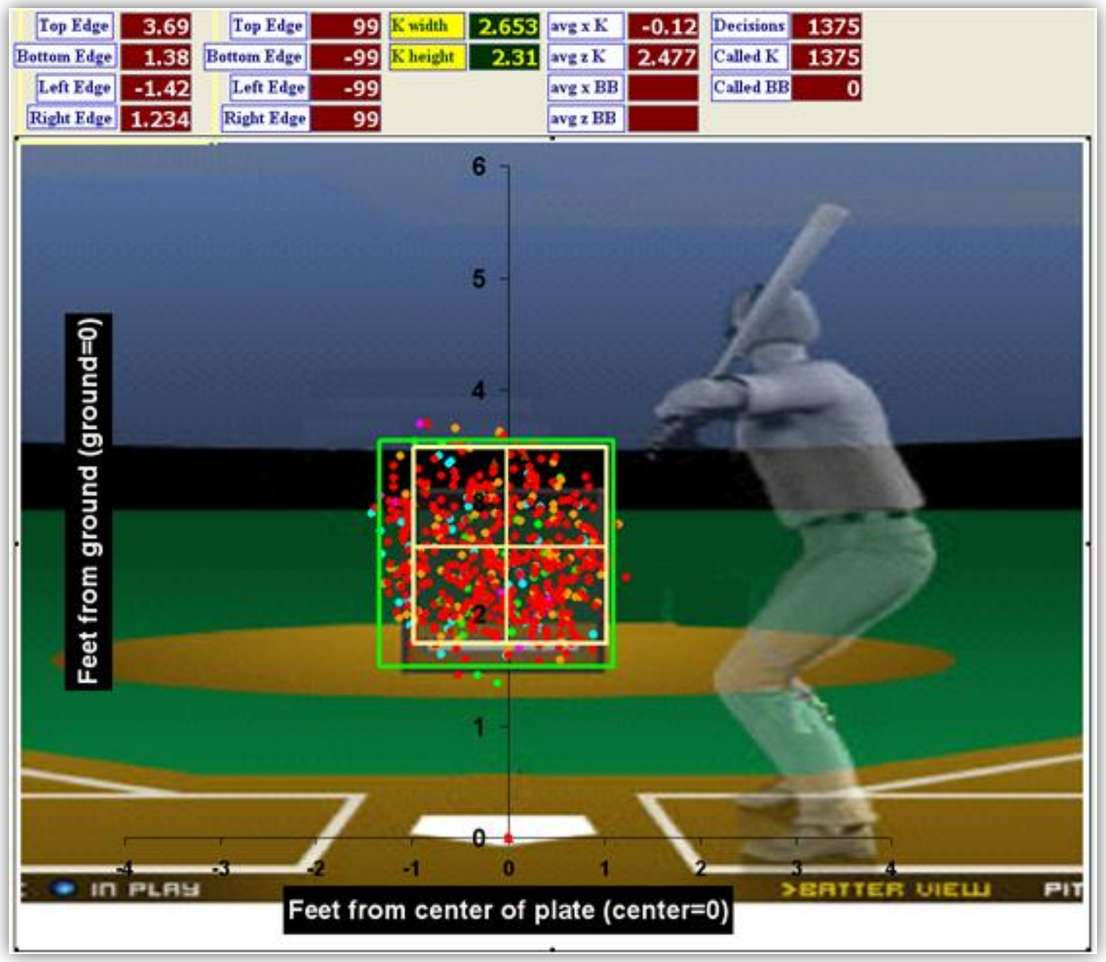


Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

The “Human Element”: Quantifying the subjective

Umpire Strike Zones

Which umpires made the most frequent bad calls?



What was Hunter Wendelstedt’s strike zone?

Umpire	% overall missed	Total calls	Balls called	Missed ball calls	Strikes called	Missed strike calls
Mike Reilly	6.71795	1950	1273	90	677	41
Eric Cooper	6.64613	2272	1481	115	791	36
Derryl Cousins	6.60935	1861	1264	91	597	32
Casey Moser	6.59341	182	127	10	55	2
Charlie Reliford	6.53144	2465	1656	80	809	81
Jim Wolf	6.53114	2312	1525	100	787	51
Mike DiMuro	6.51571	2164	1435	91	729	50
Jim Joyce	6.50919	1905	1259	78	646	46
Mike Muchlinski	6.49351	308	196	6	112	14
Tim McClelland	6.42796	2318	1591	109	727	40
	6.40632	2279	1537	101	742	45
Jeff Kellogg	6.35246	2440	1663	107	777	48
Dan Iassogna	6.27490	2008	1317	66	691	60
CB Bucknor	6.26884	1659	1082	74	577	30
Phil Cuzzi	6.01141	2279	1526	96	753	41
Kevin Causey	5.97826	184	121	7	63	4
D.J. Reyburn	5.67010	194	135	9	59	2
Mark Carlson	4.96894	161	114	5	47	3
Chris Tiller	3.90071	564	383	15	181	7

Combining Subjective and Analytic: Misplays

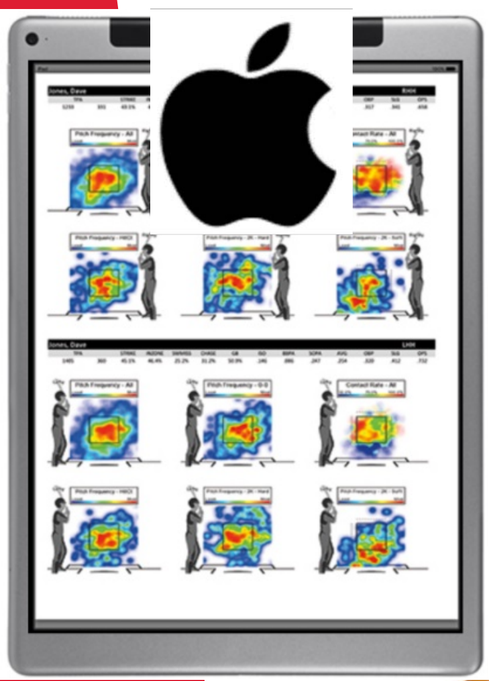
DEFENSIVE MISPLAYS: LOG

Result	Date	Inning	Batter	Fielding Pos
Failing to reach a pop foul	4/1/2016	6	Cruz, Nelson	3B
Failing to make the tag	4/5/2016	4	Braun, Ryan	3B
Ground ball through infielder	4/11/2016	5	Kershaw, Clayton E	3B
Ball stuck in glove	4/12/2016	2	Loney, James	3B
Bad throw	4/13/2016	4	Jay, Jon	3B
Losing the Double Play (Lead man)	4/27/2016	4	Utley, Chase	3B
Ground ball through infielder	4/27/2016	7	Alvarez, Pedro	3B
Failing to catch a line drive	6/19/2016	8	Weeks, Jemile	3B
Mishandling ball after safe hit	6/21/2016	1	Ruiz, Carlos	3B
Failing to catch the throw	7/3/2016	7	Turner, Justin	1B
Slow to recover	7/4/2016	8	Maybin, Cameron K	3B
Ground ball through infielder	7/4/2016	8	Maybin, Cameron K	3B
Losing the Double Play (Lead man)	7/22/2016	2	McGehee, Casey M	3B

The Future

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Movement to Mainstream



Bloomberg SPORTS

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

BIG DATA & AI LANDSCAPE 2018

INFRASTRUCTURE

HADOOP ON-PREMISE
cloudera, Hortonworks, MAPR, Pivotal, IBM InfoSphere, bluedata, jethro

HADOOP IN THE CLOUD
aws, Microsoft Azure, Google Cloud, IBM InfoSphere BigInsights, TREASURE DATA, Qubole, altiscale, CAZENA, CenturyLink

STREAMING / IN-MEMORY
aws, databricks, striim, confluent, GridGain, ORACLE Coherence, dataArtisans, hazelcast, TERRACOTTA, kx, FASTDATA, Wallaroo LABS

NoSQL DATABASES
Google Cloud, aws, ORACLE, Microsoft Azure, mongoDB, MarkLogic, AEROSPIKE, DATASTAX, ArangoDB, Couchbase, redislabs, SCYLLA

NewSQL DATABASES
SAP HANA, Clustrix, Pivotal, NUODB, Cockroach LABS, Cloud Spanner, MEMSQL, influxdata, MariaDB, TIMESCALE, VOLTDB, citusdata, splice machine, paradigm4, Trafodion, TIDB

GRAPH DBs
neo4j, Amazon Neptune, IBM, ORACLE, OrientDB, DATASTAX, InfiniteGraph, Objectivity

MPP DBs
TERADATA, VERTICA, IBM Data Warehouse Systems, Actian, Kognitio, Exasol, dremio

CLOUD EDW
aws, Google Cloud, Microsoft Azure, Pivotal, snowflake

DATA TRANSFORMATION
talend, pentaho, alteryx, TRIFACTA, tamr, Paxata, StreamSets, UNIFI

DATA INTEGRATION
SAP Data Services, Informatica, MuleSoft, TEALIUM, snapLogic, enigma, Segment, podium data, alooma, xplenty, ZALONI, Stitch, import.io, Infoworks, ATTUNITY

DATA GOVERNANCE
Informatica, IBM, SailPoint, McAfee Skyhigh Security Cloud, collibra, Alation, Waterline Data, IMMUTA, OKERA

MGMT / MONITORING
aws, New Relic, actifio, rubrik, APPDYNAMICS, DATADOG, dynatrace, WAVEFRONT by vmware, splunk>, SignalFx, druva, Moogsoft, unravel, pagerduty, Numerify, Acodex

STORAGE
aws, Google Cloud, Microsoft Azure, IBM Storage, PURESTORAGE, ALLUXIO, nimblestorage, Qumulo, panasas, COHESITY

CLUSTER SVCS
aws, kubernetes, docker, MESOSPHERE, CoreOS, pepperdata

APP DEV
Lightbend, Keen IO, rainforest, CASK

CROWD-SOURCING
amazon mechanicalturk, upwork, appen, figure eight, scale, HIVE

HARDWARE
Google TPU, arm, intel AI, GRAPHCORE, IBM Power Systems, MYTHIC, NVIDIA, cerebras, Movidius, WAVE COMPUTING, HAILO

GPU DBs
kinetica, MAPD, SQREAM, blazegraph, BLAZINGDB, brytlyt, PG-Strom

ANALYTICS

DATA ANALYST PLATFORMS
Microsoft, pentaho, alteryx, Digital Reasoning, GUAVUS, AYASDI, ATTIVIO, Datameer, Quid, incorta, interana, ClearStory, Origami logic, ENDOR, MODE, Bottlenose, switchboard

DATA SCIENCE PLATFORMS
IBM, KNIME, dataiku, DOMINO, rapidminer, CONTINUUM ANALYTICS, ALGORITHMIA, DATAWATCH ANGOSS, SAS

BI PLATFORMS
Microsoft, aws, DOMO, Wave Analytics, looker, THOUGHTSPOT, ATSCALE, ARCADIA DATA, Information Builders, SISENSE, GoodData, MicroStrategy, birst

VISUALIZATION
+tableau, SAP Lumira, Google Cloud, celonis, Qlik, Periscope Data, ZEPL, ZOOMDATA, plotly, CHARTIO, TOUCAN TOCO

MACHINE LEARNING
Azure Machine Learning, aws, Google Cloud, H2O.ai, DataRobot, gamalon, ELEMENT AI, VERSIVE, ViSENZE, deepsense.io, bonsai

COMPUTER VISION
Microsoft Azure, Amazon Rekognition, Cloud Vision API, clarifai, 商汤, EVER AI, deepomatic, twenty bn, neurala

HORIZONTAL AI
IBM Watson, Cortana, Facebook AI, sentient, Voyager Labs, Wolfram|Alpha, :) Affectiva, PROPHESEE, Cognitive Scale, Numenta, PETUUM, SI, SCALED INFERENCE, narrative logics, CURIOUS AI, OSARO, BLUE VISION

SPEECH & NLP
Google Cloud, twilio, amazon alexa, narrative science, semantic machines, Wolfram|Alpha, Mobvoi, Eigen Technologies, PRIMER, SoundHound Inc., voicea, NUANCE, MindMeld, spoken.io, snips, Speaktoit, VSAPP

SEARCH
elasticsearch, ORACLE ENDECA, EXALEAD, coveo, Lucidworks, ATTIVIO, swiftype, algolia, alphasense, MAANA, omni:us, SINEQUA

LOG ANALYTICS
splunk>, sumologic, LOGGLY, TIMBR, kibana, logz.io

SOCIAL ANALYTICS
Hootsuite, sprinklr, NETBASE, synthesio, tracx, simplereach, bitly, predata, SimilarWeb

WEB / MOBILE / COMMERCE ANALYTICS
Google Analytics, mixpanel, AMPLITUDE, sumall, Airtable, RE§CI, SIGOPT, granify, custora

APPLICATIONS – ENTERPRISE

SALES
einstein, CHORUS, INSIDESALES.COM, conversica, GONG, clari, aviso, tact.ai, fuse machines, TROOPS

MARKETING - B2B
RADIUS, App Annie, EVERSTRING, Lattice, MINTIGO, 6sense, tubular, DataFox, Reflektion, ENGAGIO, mrp

MARKETING - B2C
zeta, bloomreach, SendGrid, BlueYonder [PERSADO], kahuna, ACTIONIQ, SAILTHRU, BLUECORE, QUANTIFIND, mparticle, Amplero, amperity, TEALIUM, Simon, Lytics

CUSTOMER SERVICE
MEDALLIA, zendesk, CLARABRIDGE, Gainsight, NG DATA, DigitalGenius, afiniti, automat, frame.ai, msg.ai, INTERCOM, CaliDesk

HUMAN CAPITAL
Hire Vue, entelo, hiQ, GIGSTER, textio, RESTLESS BANDIT, Wade&Wendy, Stella, Clustree, pymetrics, mya, uncommon

LEGAL
RAVEL, Seal, Everlaw, JUDICATA, EBREVIA, IRONCLAD, PREMONITION, ROSS, casetext

FINANCE
Anaplan, zuora, SAP S/4 HANA, TRADESHIFT

ENTERPRISE PRODUCTIVITY
slack, ORACLE, X., lumiata, DIFFBOT, clara, talla, butter.ai, Kasisto

BACK OFFICE AUTOMATION
UiPath, HyperScience, blueprism, Kofax, Contrixity, AppZen, WorkFusion

SECURITY
TANIUM, CYLANCE, zscaler, StackPath, illumio, CODE42, CipherCloud, DARKTRACE, ANOMALI, ThreatMetrix, VECTRA, cybereason, Guardian Analytics, DATAVISOR, siftscience, SIGNIFYD, SentinelOne, SecurityScorecard, SOCURE, BlueTalon, Recorded Future, feedzai, Cybereason, ARFA1 SECURITY, sparkcognition, IronNet Cybersecurity

APPLICATIONS – INDUSTRY

ADVERTISING
AppNexus, MediaMath, criteo, AdRoll, Integral Ad Science, ORACLE MOAT, OpenX, datorama, theTradeDesk, Algorithms, dstillery, LiveIntent, TAPAD, dataxu, gumgum, Qppier, DYNAMIC YIELD, teemo, yieldmo

EDUCATION
Liulishuo, KNEWTON, Clever, Declara, kidaptive, PANORAMA, knowre, gradescope

GOVERNMENT
OPENGOV, mark43, FiscalNote, GRIDSMART, LiveStories, Passport, SmartProcure, STREETLIGHTDATA, OpenDataSoft

REAL ESTATE
REDFIN, Opendoor, VTS, CREDIFI, reonomy, COMPSTAK, CAPE

FINANCE - INVESTING
KENSHO, Dataminr, Quantopian, ADDEPAR, NUMERAI, iSENTIUM, ALGORIZ, RavenPack, AYASDI

FINANCE - LENDING
ondeck, Affirm, 拍拍贷, JIANPU.AI, pedal.com, Kreditech, AVANT, TALA, zestfinance, Upstart, INSIKT, upgrade, 100Credit, WeLab, Wecashi财富科, TrueAccord, MoneyLion, Active.AI, aire, cignifi

INSURANCE
Metromile, Lemonade, CYENCE, Shift Technology, TRACTABLE

HEALTHCARE
flatiron, Clover, KYRUUS, HealthTap, METABIOTA, Ginger.io, Glow, babylon, 3DMed, zebra, PathAI, OVIA, TEMPUS, patientslikeme, AiCure, RECURSION, prognos, enlitic, imagia, Qventus, BAYLABS, ARTERYS, CLOUDMEDX, IMAGEN, Kang Health, PAIGE, DATAVANT, HUMAN DIAGNOSIS PROJECT, innoviccer, JOONTAQS

LIFE SCIENCES
23andMe, color, Genapsys, BenevolentAI, verily, WuXiNextCODE, ZEPHYR HEALTH, ZS, SCAN, Clear Labs, freenome, NANOPORE, Phosphorus, DNAnexus, CITRINE, twoXAR, Atomwise, deep genomics, SOPHiA, OWKIN

TRANSPORTATION
UBER, TESLA, ZOOX, CLEARPATH, nuTonomy, drive.ai, nauto, AUTOMOTIVE, PILOT.AI, NIO, OPTIMUS RIDE, moovit, nexar, netradyne, comma.ai, Civil Maps, Cruise Automation

AGRICULTURE
FARMERS EDGE, Granular, JOHN DEERE, BLUE RIVER TECHNOLOGY, FarmersEdge, FarmLogs, TARANIS, GAMAYA, TerrAvion, prospera

COMMERCE
Instacart, STITCH FIX, Dia & Co, RetailNext, HowGood, heuritech

INDUSTRIAL
AVEVA, SIEMENS, Schneider Electric, PREDIX, GE IIoT, UPTAKE, OSIsoft, SIGHT MACHINE, TACHYUS, Alluvium, SCORTEX

OTHER
charmony, stem, rethink robotics, Amper, ByteDance, happer, celect, BOXEVER, VERDIGRIS, duetto, Unbabel, Jukedeck, Skycatch, Spectrum, remesh, ASAPP

CROSS-INFRASTRUCTURE/ANALYTICS
aws, Google Cloud, Microsoft, IBM, SAP, Hewlett Packard Enterprise, SAS, 1010DATA, vmware, TIBCO, TERADATA, ORACLE, NetApp, syncsort, MAPR, cloudera

OPEN SOURCE

FRAMEWORK
hadoop HDFS, hadoop MapReduce, Apache YARN, Flink, TEZ, Spark, MESOS, CDAP

QUERY / DATA FLOW
Spark SQL, presto, HIVE, SLAMDATA, APACHE DRILL, Google Cloud Dataflow, Flink

DATA ACCESS
cassandra, nifi, mongoDB, CouchDB, OpenTSDB, SciDB, riak, APACHE HBASE, Cloud Spanner, ACCUMULO

COORDINATION
talend, Apache Zookeeper, CHUBBY, Apache Ambari, Apache Airflow

STREAMING
Spark, APEX, Flink, beam, kafka, druid, STORM

STAT TOOLS
python, R, ScalaLab, NumPy, SciPy, julia

AI / MACHINE LEARNING / DEEP LEARNING
TensorFlow, theano, torch, MADlib, learn, Caffe, Microsoft Cognitive Toolkit, OpenAI, DMTK, Keras, PaddlePaddle, Apache SINGA, FeatureFu, mxnet, Chainer, VELES, WEKA, neon™, DIMSUM, DSSTNE, mllib, DL4J, MAHOUT, Aerosolve

SEARCH
elasticsearch, Solr, Lucene

LOGGING & MONITORING
elasticsearch, kibana, SENTRY, logstash, Prometheus

VISUALIZATION
BeakerX, Rodeo

COLLABORATION
jupyter, Zeppelin, ANACONDA

SECURITY
Apache Ranger, KNOX, Sentry

DATA SOURCES & APIs

HEALTH
VALIDIC, practicefusion, fitbit, GARMIN, HUMAN API, Qsimplicity

IOT
GE Digital, UPTAKE, thingworx, helium, samsara, AUGURY, estimote

FINANCIAL & ECONOMIC DATA
Bloomberg, THOMSON REUTERS, DOW JONES, S&P CAPITAL IQ, CBINSIGHTS, xignite, Quandl, ENVESTNET YODLEE, PREMISE, estimize, SECOND MEASURE, Eagle Alpha, StockTwits, PLAID, Thinknum, earnest

AIR / SPACE / SEA
Orbital Insight, planet., SKYCATCH, Airware, AIROBOTICS, spire, kespry, PRECISIONHAWK, UNDERSTORY, Descartes Labs, WINDWARD, tellusLabs, DroneDeploy, MarineTraffic

PEOPLE / ENTITIES
acxiom, experian, EPSILON, InsideView, Crimson Hexagon, BASIS TECHNOLOGY, Quantcast, SAFEGRAPH

LOCATION INTELLIGENCE
FOURSQUARE, MapAnything, mapbox, sense360, pitney bowes, HEXAGON, PlaceIQ, esri, factual, CARTO, Mapillary, StreetLine, cuebiq, Radar

OTHER
qualtrics, DATA.GOV, data.world, enigma, mobilewalla, CRUX, HIGHCHARTS

DATA RESOURCES

DATA SERVICES
Palantir, UC Berkeley RISELab, OPERA, MU SIGMA, DATA SCIENCE, fractal, kaggle, DataKind, EXL, INNOPLEXUS

INCUBATORS & SCHOOLS
PLURALSIGHT, GA, galvanize, DataCamp, DataElite, INSIGHT, The Data Incubator, METIS

RESEARCH
OpenAI, facebook research, MILA, MIRI, VECTOR INSTITUTE, CSAIL, DFKI, AI2, ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

Final 2018 version, updated 07/15/2018

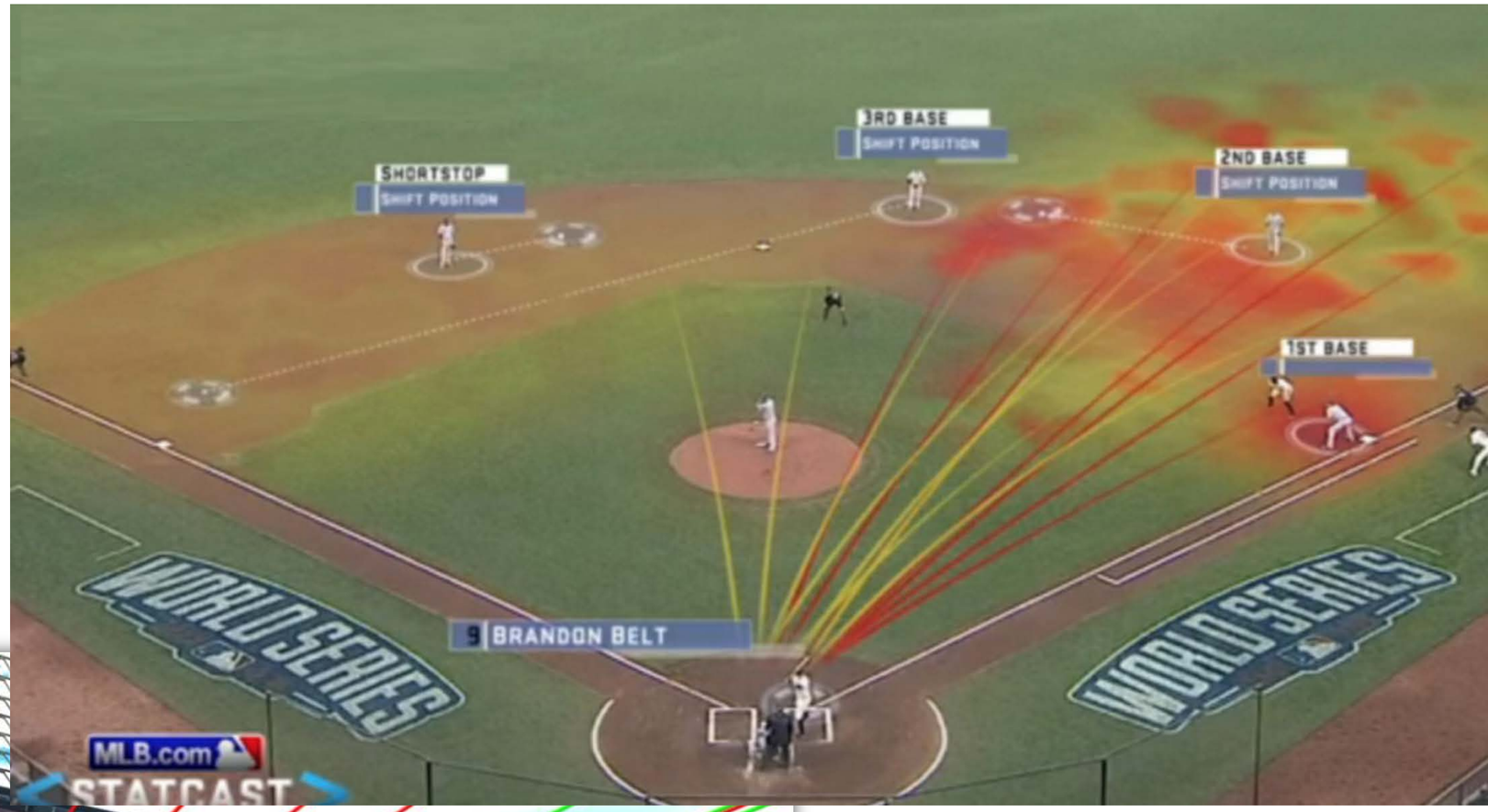
© Matt Turck (@mattturck), Demi Obayomi (@demi_obayomi), & FirstMark (@firstmarkcap)

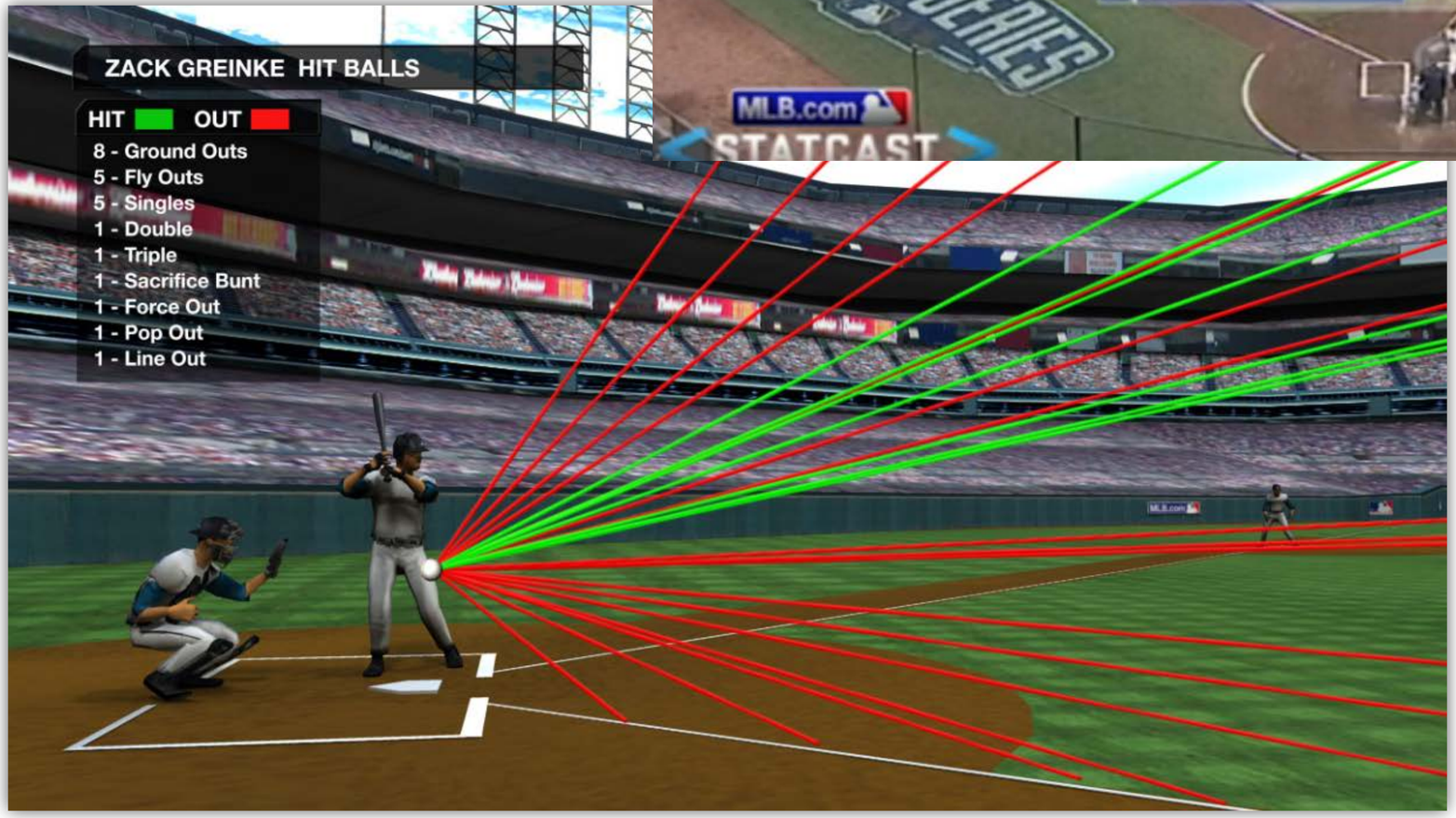
mattturck.com/bigdata2018

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

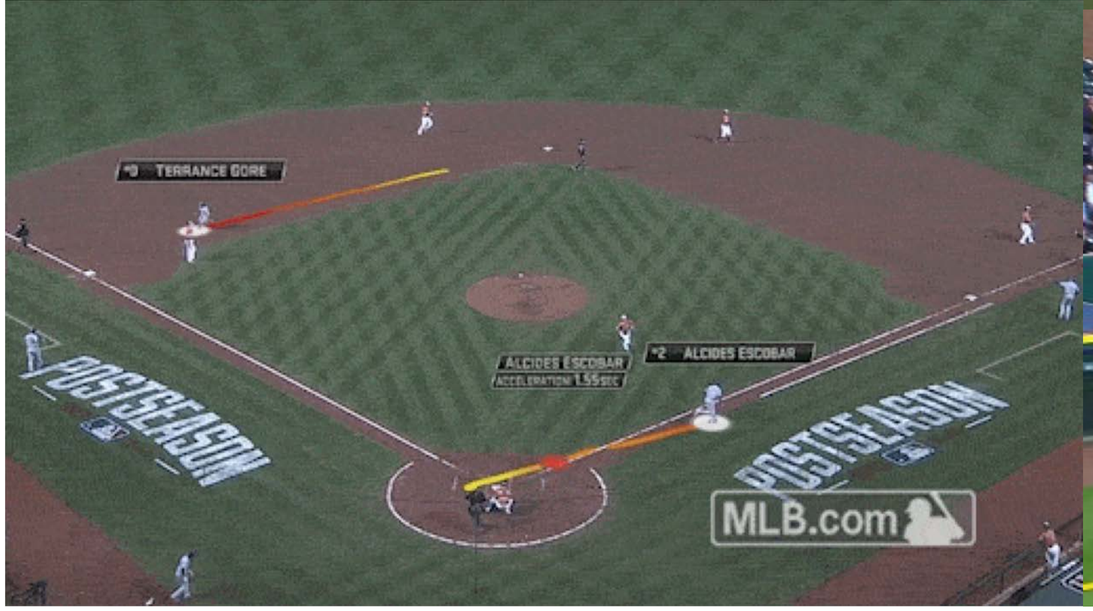
The future of sports analysis

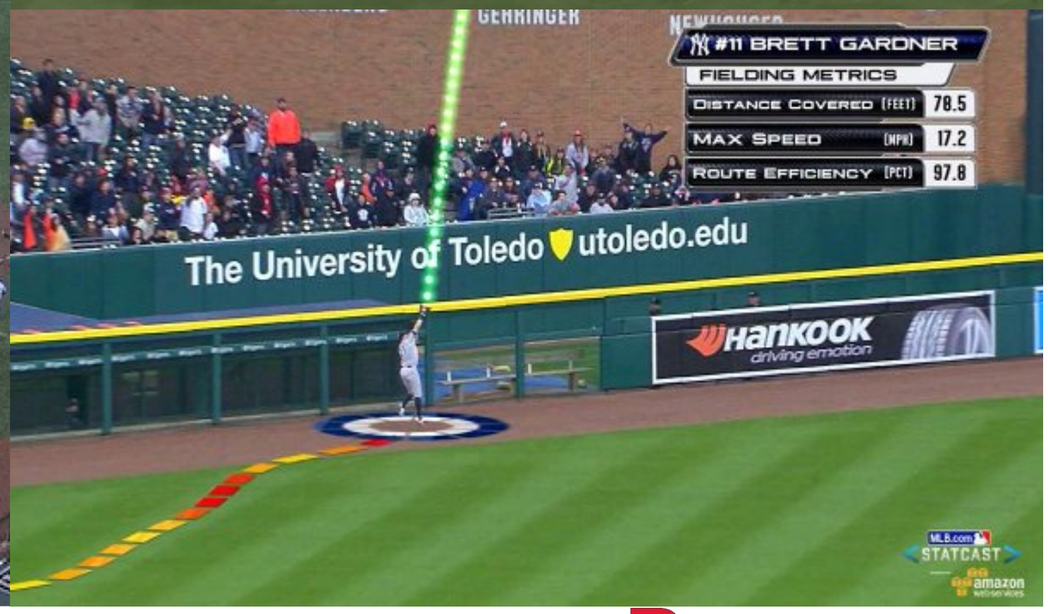




Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Create the ULTIMATE Game Plan

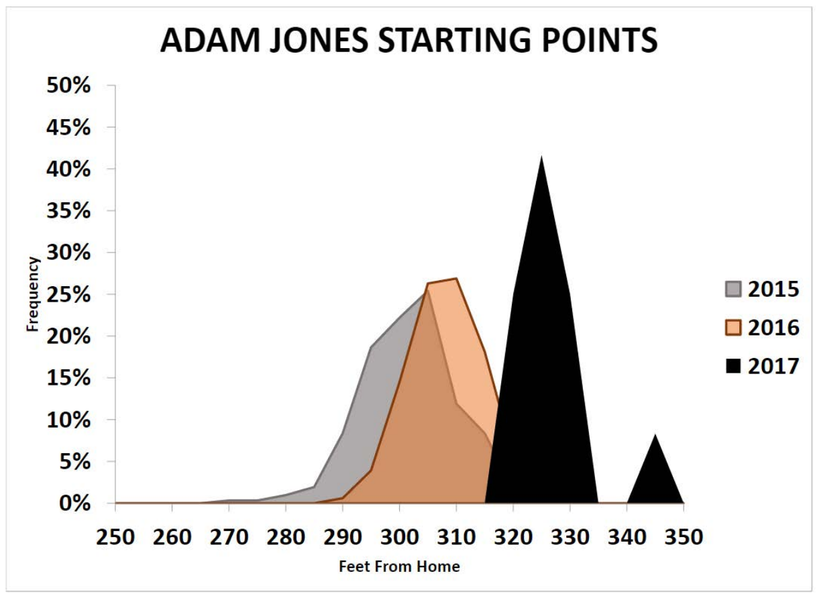




Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Example of Success: Adam Jones

Adam Jones move back 15 feet in the outfield 2018

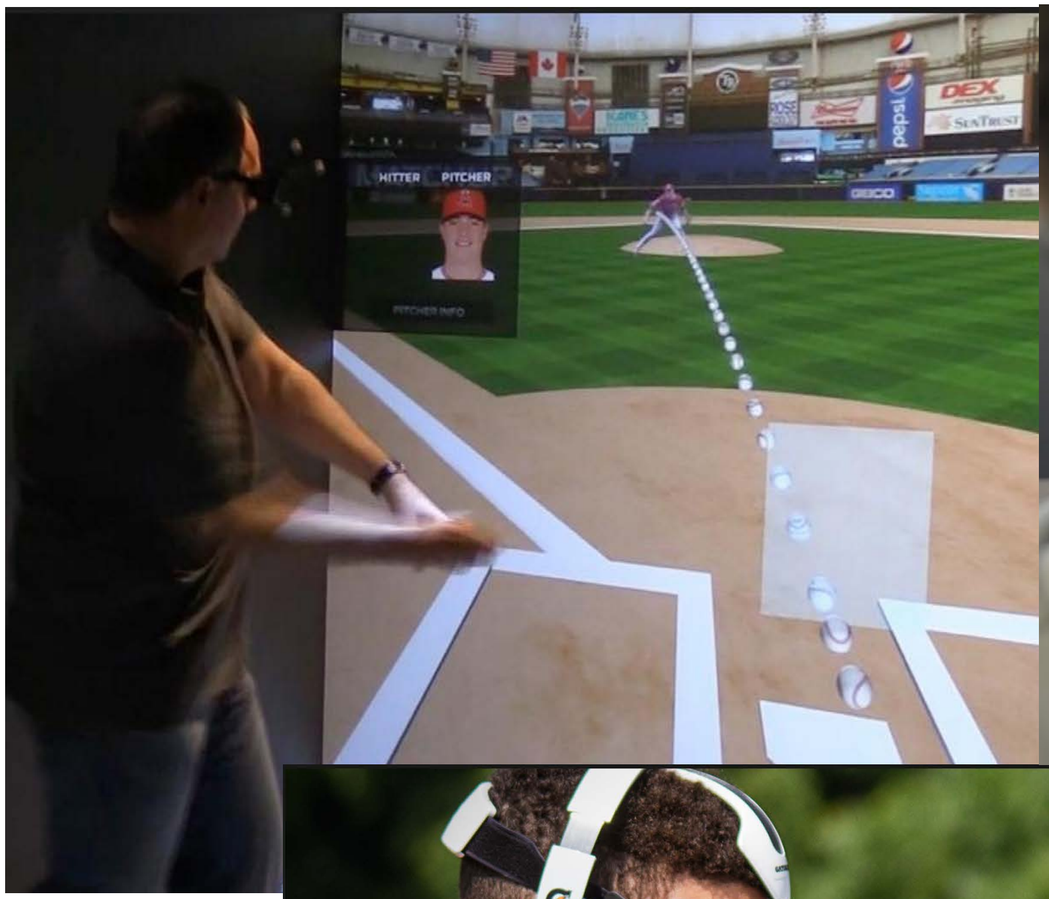


- 2017: Jones was positioned 303 feet away from home plate. Richie Shaffer crushed the ball 393 feet. Jones was unable to track it down losing the game.
- 2018: Kevin Pillar hit a similarly-angled ball a projected 398 feet away, Jones was positioned 330 feet deep. Only needed to run 68 feet vs 90 feet. Not only did Jones manage to get there, he made it look so easy that he blew a bubble while hauling it in.

"The number guys are smarter than the players," Jones told MLB.com when asked about the change this week. "It's weird playing a little deeper, but that's the way our front office wants me to play. I'm not insubordinate. I will do what they ask and sacrifice in other areas. That's what they see in the data."

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Technology: Augmented Reality



Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Thank you for attending this session!
Your bonus code is:

MLB

Go to www.RavingPlay.com to enter this code!
Powered by Engaged Nation visit booth #12 for info

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Appendix

for Q&A

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Statistical Analysis: Predicting Issues

- Injury database: http://www.spotrac.com/mlb/disabled-list/2018
- 2018 injuries cost $746 million ; 574 players ; 34,126 days on DL
- 2017 injuries cost $614 million ; 528 players ; 31,344 days on DL
- 2016 injuries cost $691 million ; 475 players ; 40,617 days on DL

Reserve/Disabled List By Team

TEAM	PLAYERS	DAYS	2016 PAYROLL
Los Angeles Dodgers	26	2,551	$82,862,281
Los Angeles Angels of Anaheim	19	1,988	$48,851,243
Texas Rangers	19	1,693	$46,128,077
New York Yankees	18	1,888	$42,321,434
New York Mets	17	1,401	$41,765,396

REASON	PLAYERS	DAYS	2016 PAYROLL
Shoulder	89	7,787	$147,179,794
Back	38	2,435	$75,698,930
Hamstring	44	2,585	$73,157,013
Knee	38	3,187	$53,000,638
Elbow	54	5,973	$50,651,053
Elbow Tommy John	21	3,222	$49,757,019

Reserve/Disabled List Bv Position

POSITION	PLAYERS	DAYS	2016 PAYROLL
Starting Pitcher	106	9,801	$264,114,970
Relief Pitcher	168	15,101	$122,327,845
3rd Base	18	1,627	$62,076,883
Left Field	34	2,740	$42,508,634
Center Field	29	2,030	$40,588,411
Designated Hitter	7	572	$40,587,720

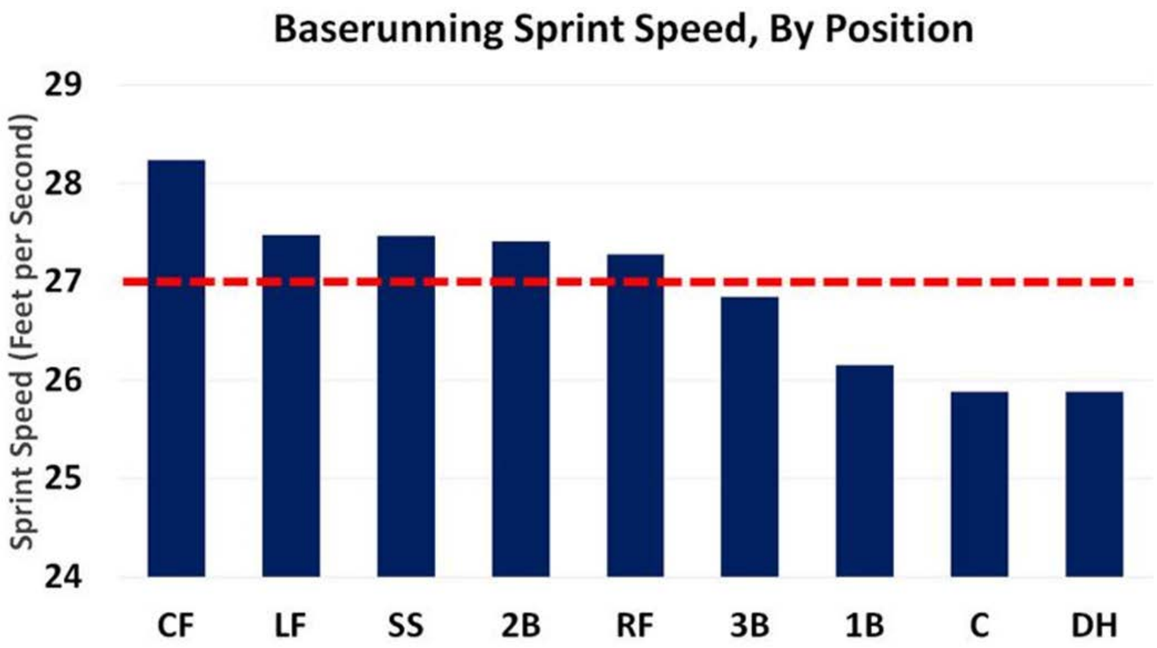
Reserve/Disabled List By Player

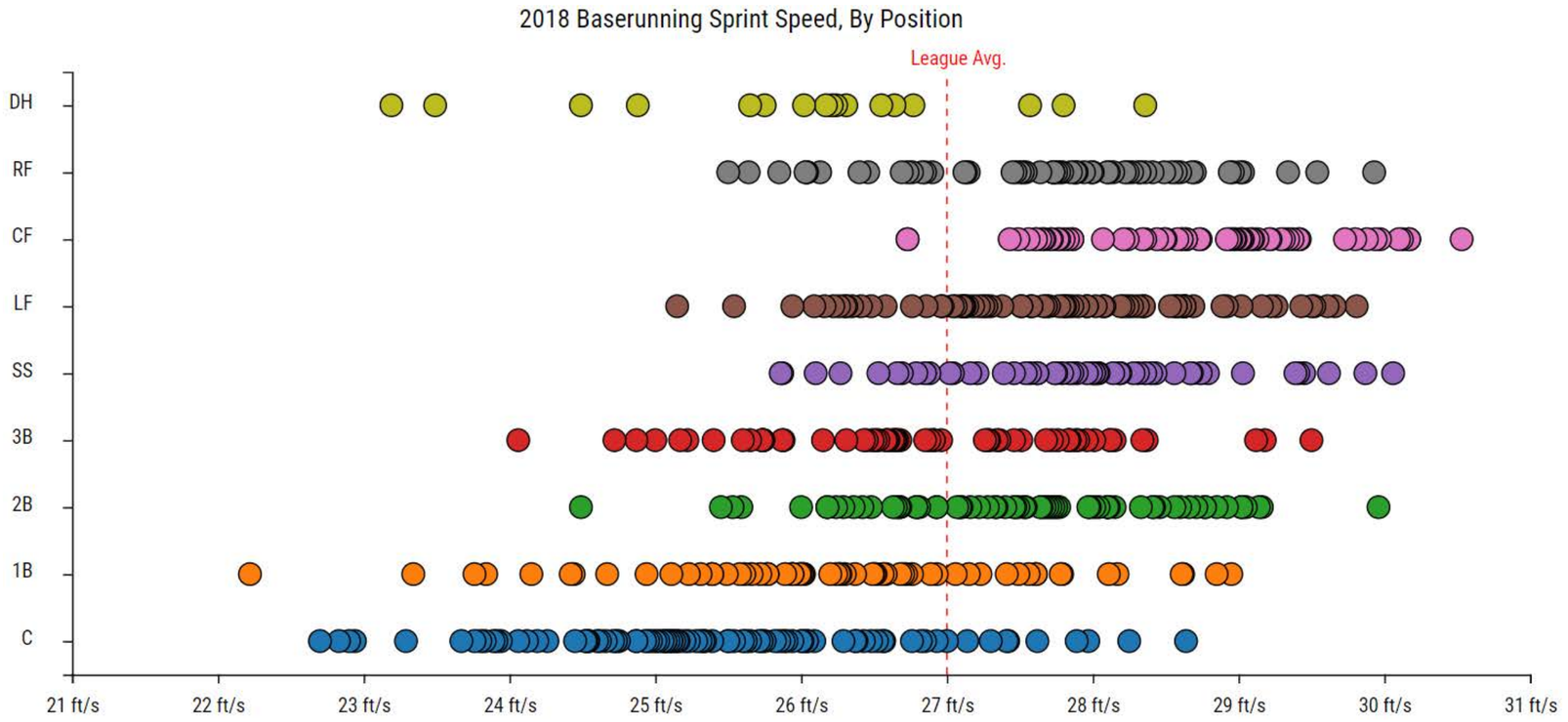
PLAYER	POS	TEAM	REASONS	DAYS	2016 PAYROLL
C.J. Wilson	SP	LAA	Shoulder	232	$25,355,280
Alex Rodriguez	DH	NYY	Hamstring	200	$21,858,000
Pablo Sandoval	3B	BOS	Shoulder	222	$21,350,850
David Wright	3B	NYM	Neck	175	$19,125,750
Matt Harrison	SP	PHI	Back	232	$16,480,816
Andre Ethier	CF	LAD	Leg	160	$15,737,760

Statistical Analysis: Predicting Injuries

	COEFFICIENTS	STANDARD ERROR	P-VALUE
Intercept	1.6319	0.27	0.00
Average of LHP?	-0.1847	0.07	0.01
Avg Arm Slot STDDEV	1.6667	0.54	0.00
Arm/Shoulder?	0.0110	0.00	0.00
Previous TJS?	0.2981	0.07	0.00
Hard Pitches	0.0001	0.00	0.15
ERA-	-0.0020	0.00	0.04
Age	-0.0524	0.01	0.00

StatCast: Baserunning Sprint Speed

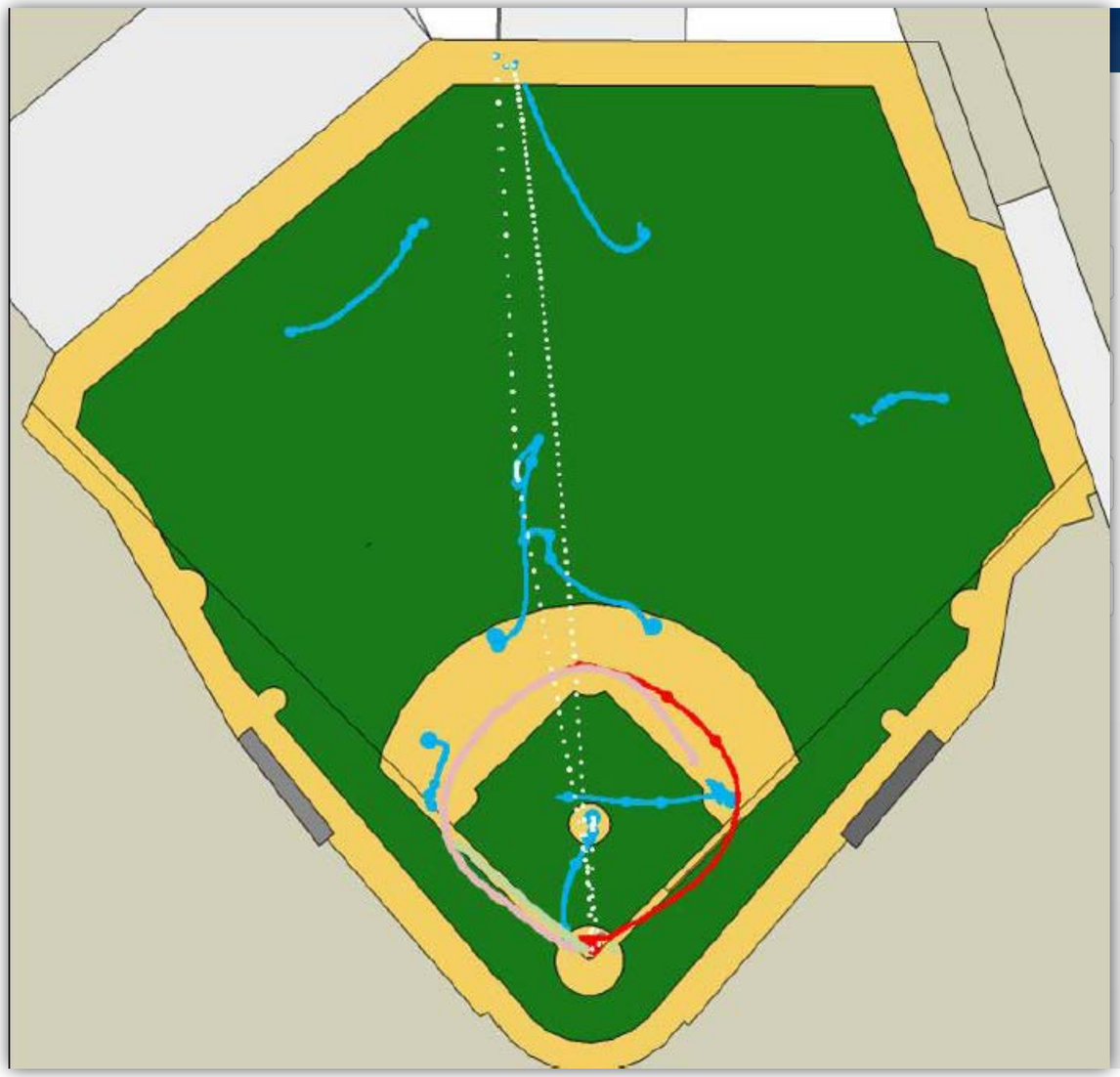




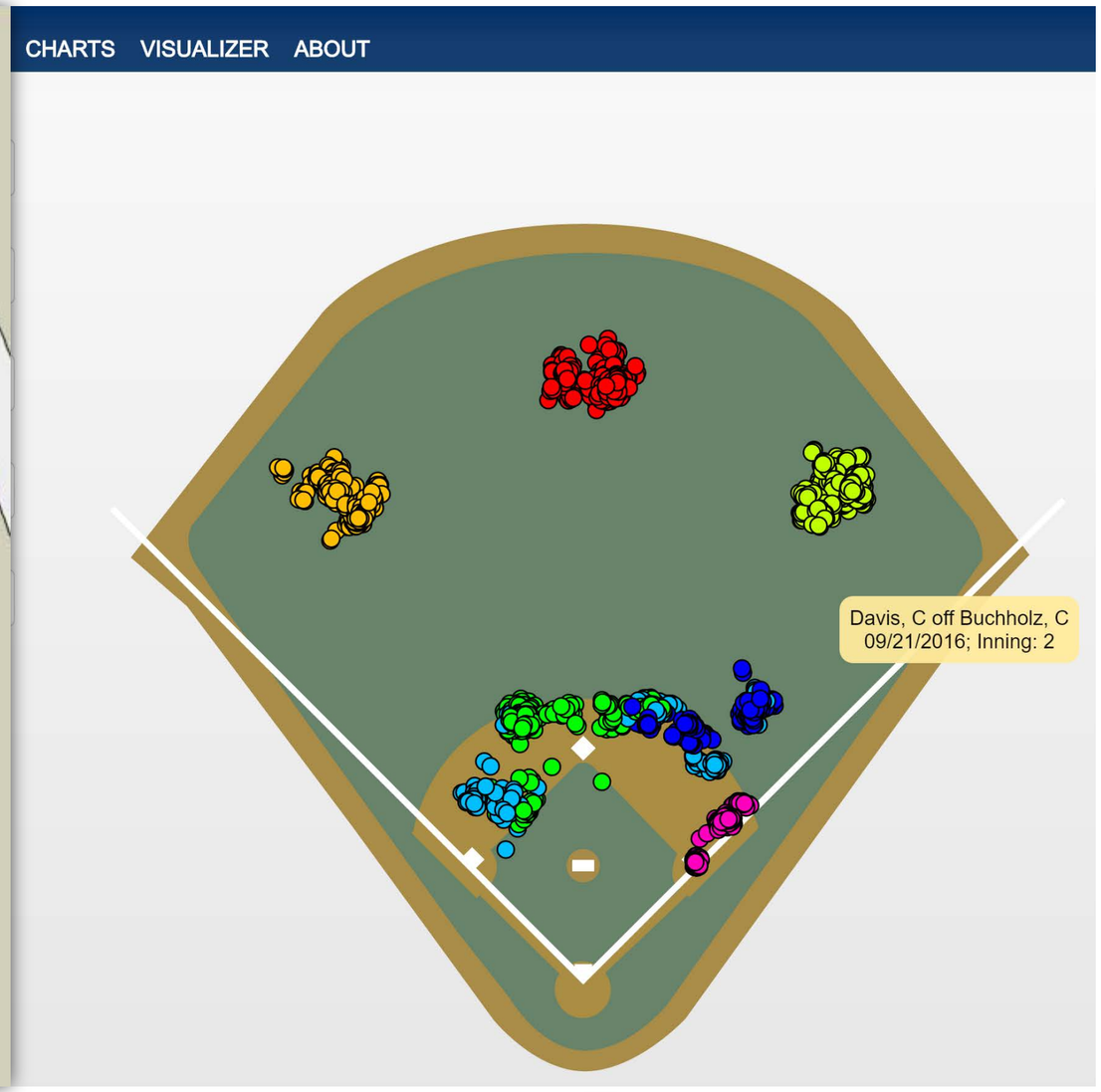
RavingNEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Create the ULTIMATE game-plan

FieldFX: Tracking every fielder



StatCast: Measuring the Shift

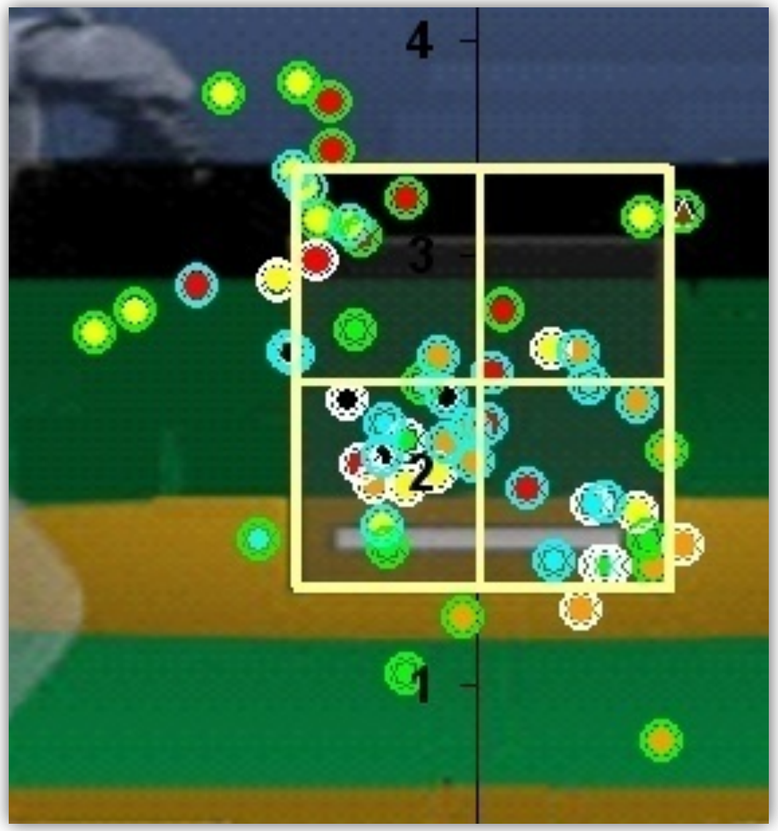


Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

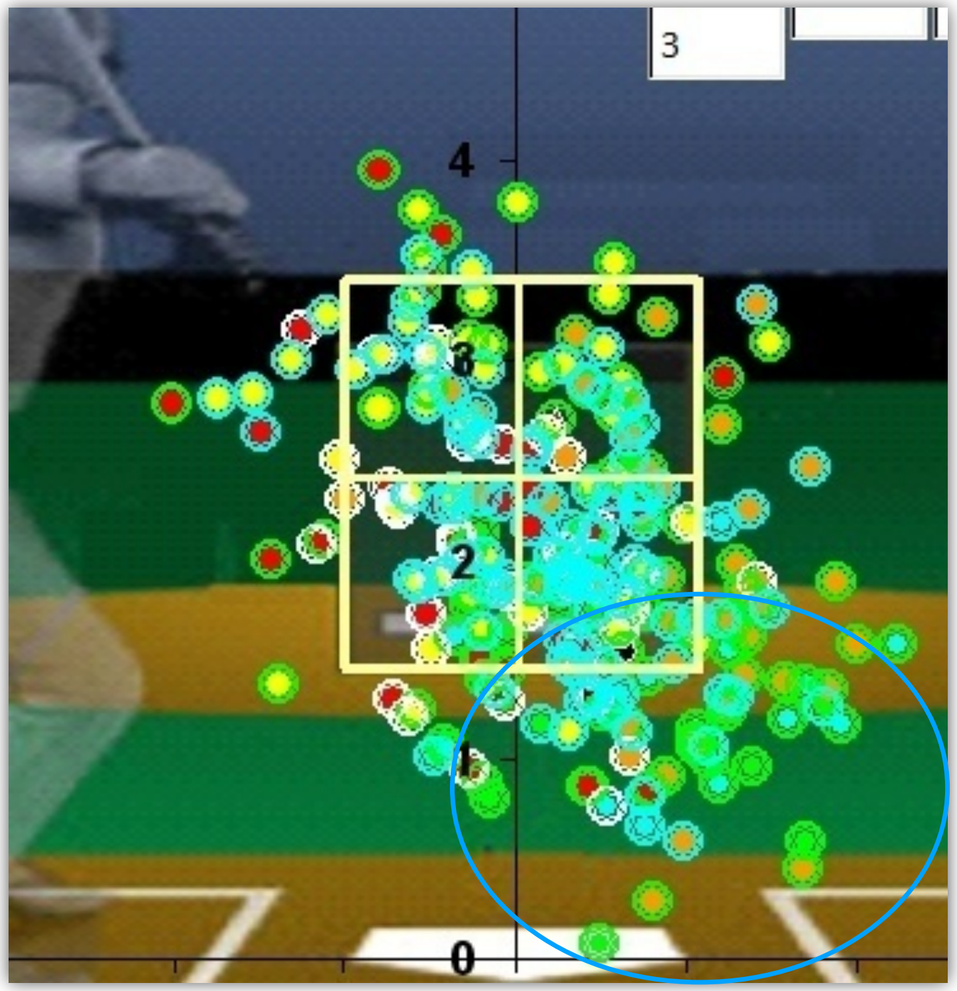
Find the “signal from the noise”

Example: is a player’s hurt knee affecting their swing?

Before injury: white circles are hits, green are misses



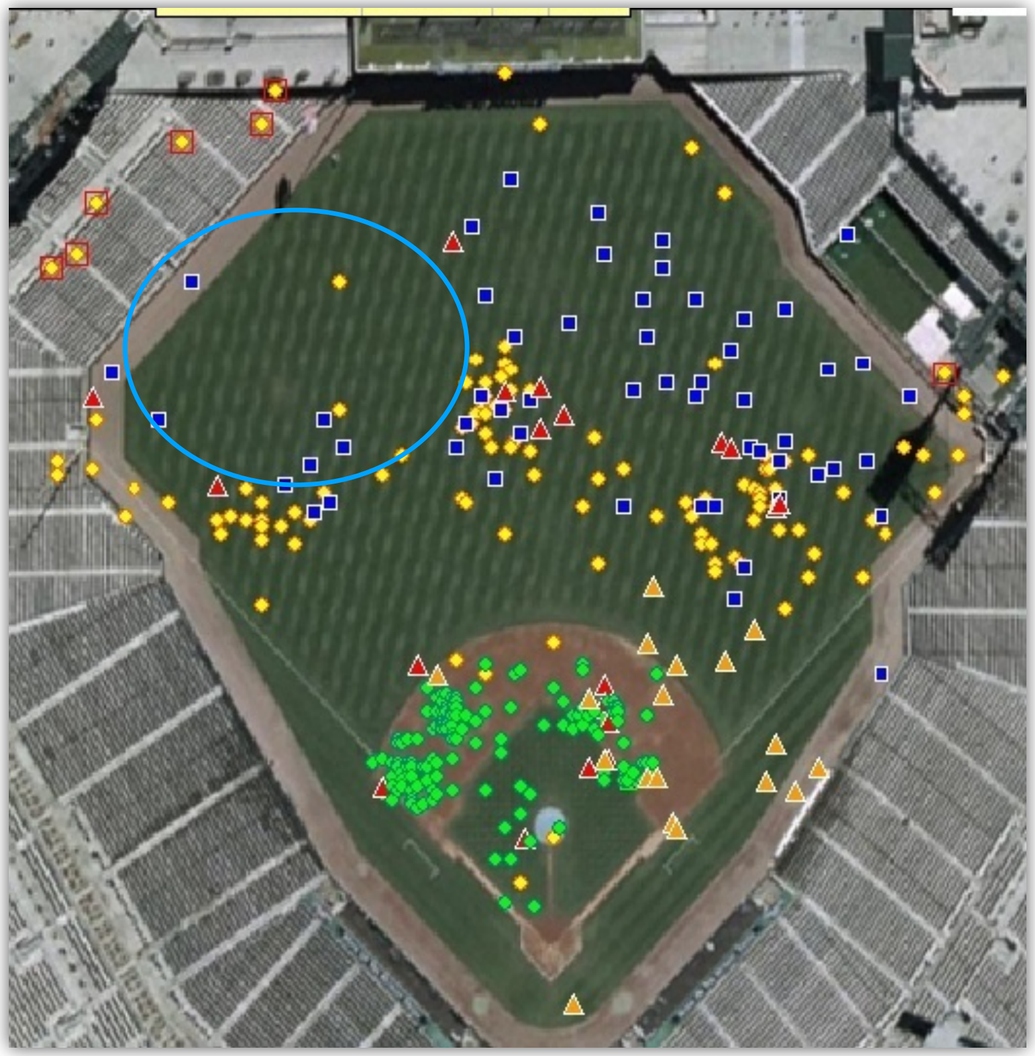
After injury

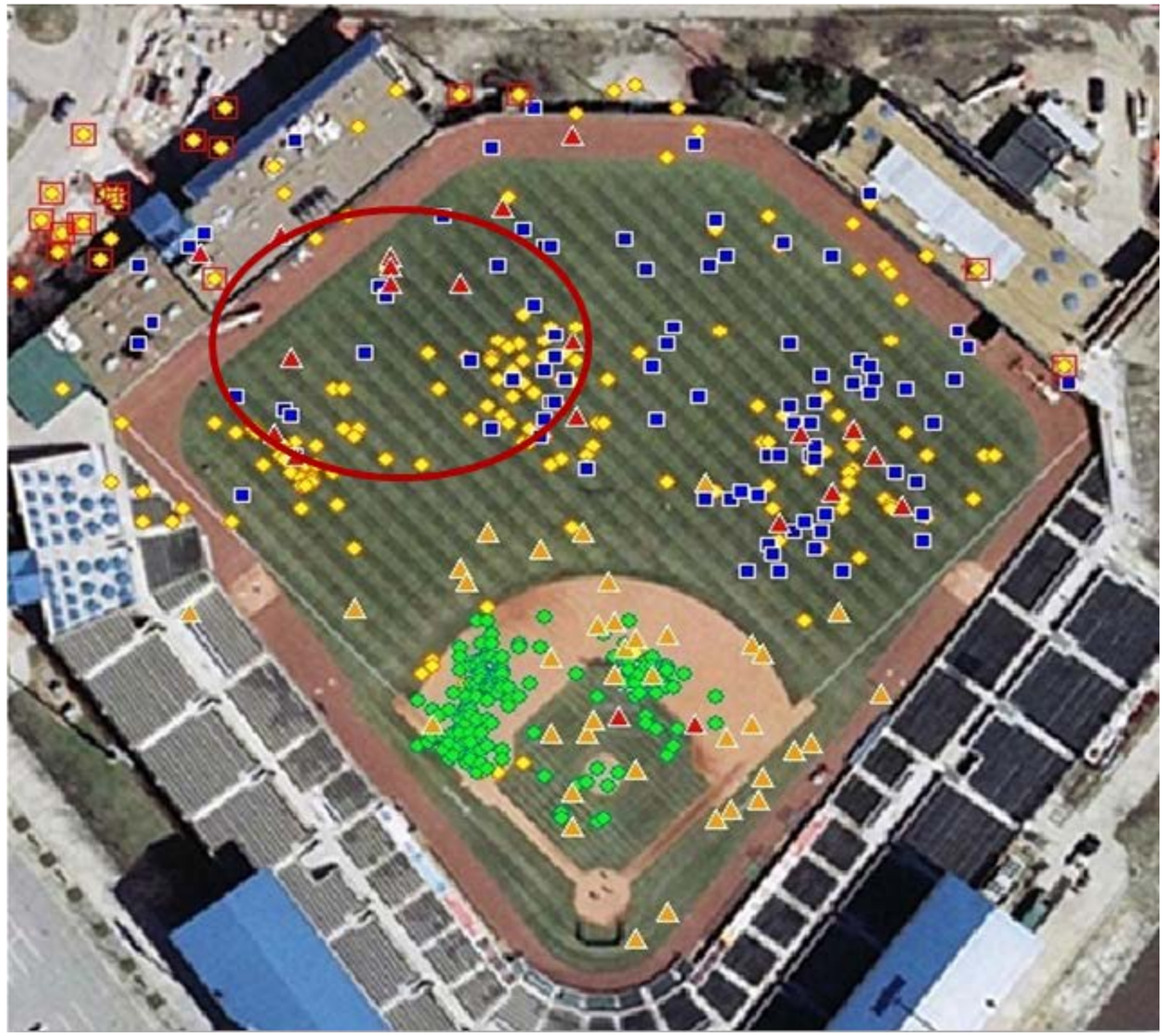


Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Use intelligence to better “defend” yourself

Be “positioned” to win





Raving NEXT

INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

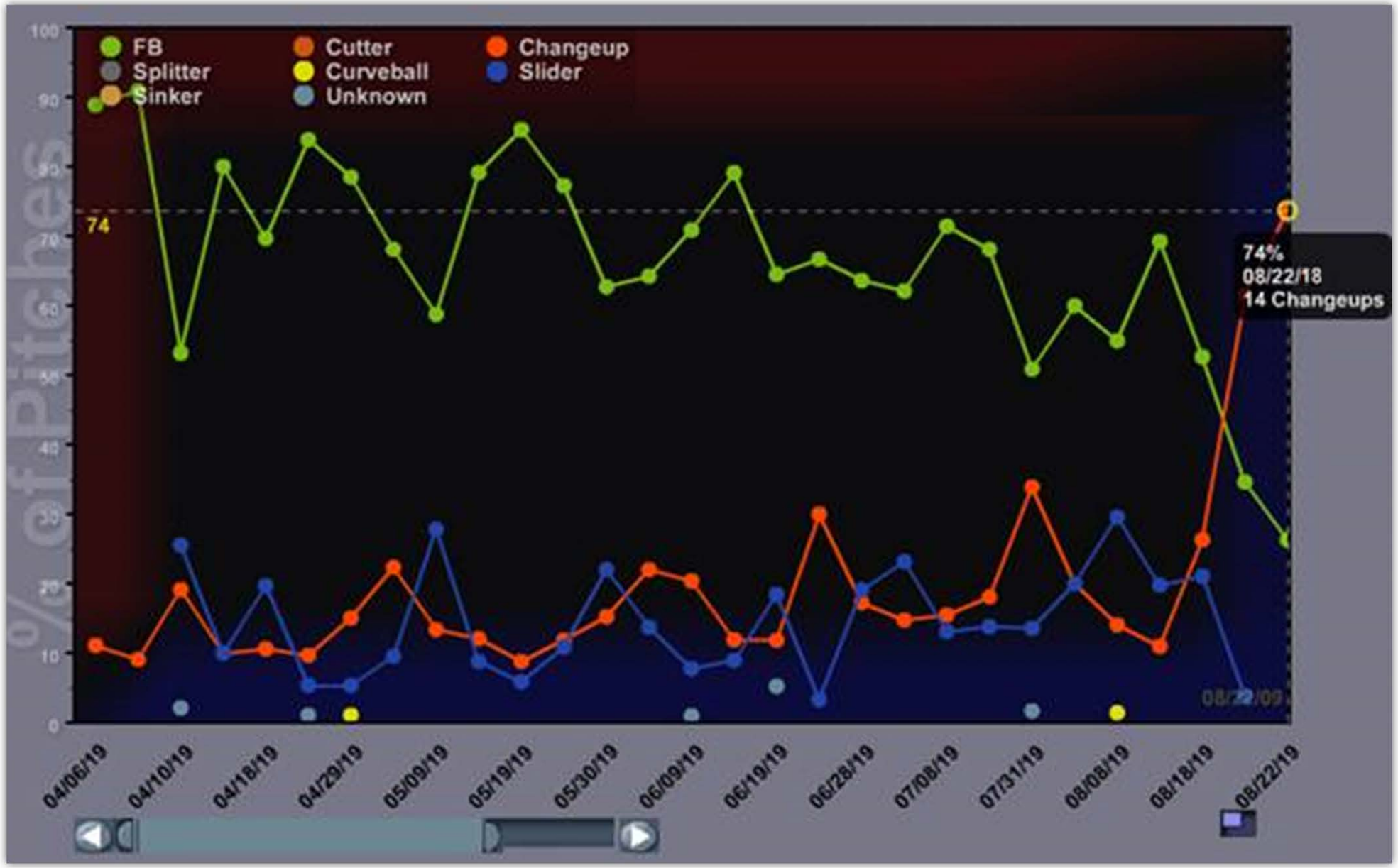
Analytics for past, present, future

Learn these three points:

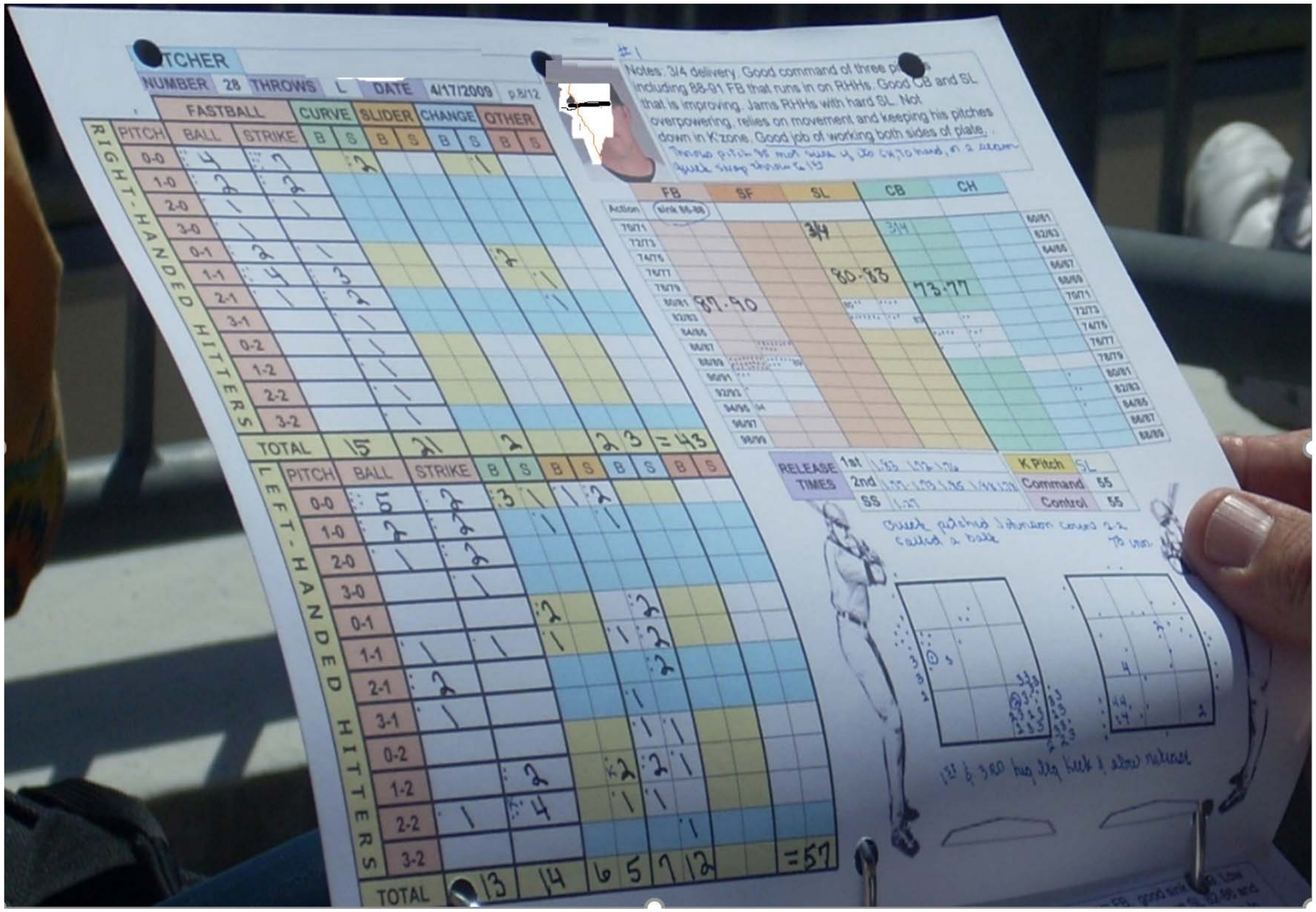
What has happened?
What is happening?
What will happen?

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Predict what might happen from time-series information

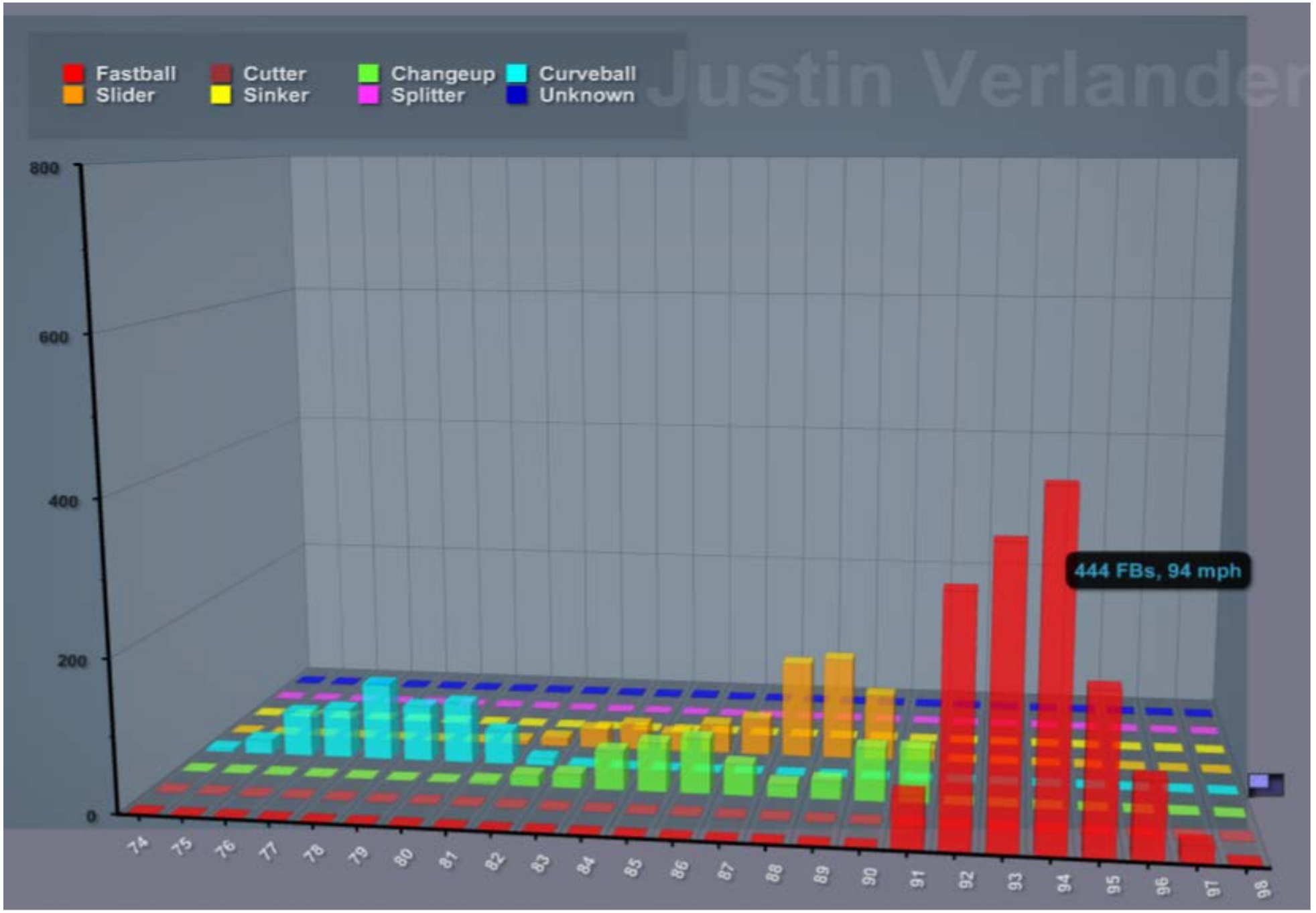


Advanced Scouting: data collection for predictive analysis



Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Statistical Analysis: Histogram



RavingNEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

The future of fielding analysis

The pivot: 6-4-3 Double plays

Play #	Time from SS to 2B*	Pivot time**
1	.60	.60
2	.40	.40
3	.27	.60
4	.53	.40
5	.53	.40
6	.53	.53
7	.67	.33

Competition: put yourself in their shoes

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Sample Data Feed


```
▼<atbat num="67" b="2" s="2" o="2" start_tfs="031021" start_tfs_zulu="2016-11-03T03:10:21Z" batter="434658" stand="R" b_height="5-10" pitcher="547973" p_throws="L" des="Rajai Davis homers (1) on a line drive
to left field. Brandon Guyer scores. " des_es="Rajai Davis batea jonrón (1) con línea por el jardín izquierdo. Brandon Guyer anota . " event_num="561" event="Home Run" event_es="Jonrón" play_guid="013f9eac-
3005-4a7c-af8b-29c5f6678c63" score="T" home_team_runs="6" away_team_runs="6">
  <pitch des="Ball" des_es="Bola mala" id="550" type="B" tfs="031032" tfs_zulu="2016-11-03T03:10:32Z" x="84.26" y="152.06" event_num="550" on_2b="446386" sv_id="161102_231341" play_guid="9becfc37-b779-4d95-
  b9d0-b0b81895d000" start_speed="98.4" end_speed="90.5" sz_top="3.37" sz_bot="1.47" pfx_x="5.35" pfx_z="12.04" px="0.859" pz="3.212" x0="1.145" y0="50.0" z0="6.136" vx0="-2.792" vy0="-144.158" vz0="-7.131"
  ax="11.264" ay="33.637" az="-6.766" break_y="23.8" break_angle="-51.6" break_length="2.6" pitch_type="FF" type_confidence=".896" zone="12" nasty="79" spin_dir="156.093" spin_rate="2797.029" cc="" mt=""/>
  <pitch des="Foul" des_es="Foul" id="551" type="S" tfs="031050" tfs_zulu="2016-11-03T03:10:50Z" x="123.48" y="186.99" event_num="551" on_2b="446386" sv_id="161102_231359" play_guid="1275d0a1-5868-4c56-
  8aa1-c463fbf51738" start_speed="97.8" end_speed="90.0" sz_top="3.3" sz_bot="1.48" pfx_x="6.07" pfx_z="12.06" px="-0.17" pz="1.918" x0="0.974" y0="50.0" z0="6.12" vx0="-5.458" vy0="-142.934" vz0="-10.603"
  ax="12.574" ay="32.58" az="-7.121" break_y="23.8" break_angle="-50.1" break_length="2.9" pitch_type="FF" type_confidence=".881" zone="8" nasty="34" spin_dir="153.349" spin_rate="2845.189" cc="" mt=""/>
  <pitch des="Foul" des_es="Foul" id="552" type="S" tfs="031118" tfs_zulu="2016-11-03T03:11:18Z" x="147.23" y="175.9" event_num="552" on_2b="446386" sv_id="161102_231428" play_guid="fba777ca-7b3c-4e2c-89df-
  e5c7a5045f05" start_speed="97.1" end_speed="89.9" sz_top="3.3" sz_bot="1.48" pfx_x="4.23" pfx_z="10.47" px="-0.793" pz="2.329" x0="0.838" y0="50.0" z0="6.148" vx0="-6.132" vy0="-141.94" vz0="-8.853"
  ax="8.693" ay="30.225" az="-10.577" break_y="23.8" break_angle="-30.8" break_length="2.9" pitch_type="FF" type_confidence=".850" zone="13" nasty="47" spin_dir="158.076" spin_rate="2379.559" cc="" mt=""/>
  <pitch des="Ball" des_es="Bola mala" id="553" type="B" tfs="031150" tfs_zulu="2016-11-03T03:11:50Z" x="58.53" y="158.73" event_num="553" on_2b="446386" sv_id="161102_231500" play_guid="abc1fd59-6ba4-4f75-
  a0ea-2f7e8596e501" start_speed="99.8" end_speed="92.4" sz_top="3.3" sz_bot="1.48" pfx_x="5.04" pfx_z="11.69" px="1.534" pz="2.965" x0="1.045" y0="50.0" z0="6.183" vx0="-0.484" vy0="-146.192" vz0="-8.173"
  ax="10.988" ay="32.177" az="-6.611" break_y="23.8" break_angle="-52.3" break_length="2.6" pitch_type="FF" type_confidence=".902" zone="12" nasty="30" spin_dir="156.741" spin_rate="2761.252" cc="" mt=""/>
  <pitch des="Foul" des_es="Foul" id="554" type="S" tfs="031210" tfs_zulu="2016-11-03T03:12:10Z" x="147.3" y="195.66" event_num="554" on_2b="446386" sv_id="161102_231520" play_guid="4282228a-48a4-4f9e-8095-
  58017d003f56" start_speed="98.1" end_speed="91.0" sz_top="3.3" sz_bot="1.48" pfx_x="3.1" pfx_z="10.98" px="-0.795" pz="1.597" x0="0.914" y0="50.0" z0="6.031" vx0="-5.998" vy0="-143.389" vz0="-11.003"
  ax="6.504" ay="30.614" az="-9.053" break_y="23.8" break_angle="-24.8" break_length="2.4" pitch_type="FF" type_confidence=".875" zone="13" nasty="61" spin_dir="164.290" spin_rate="2430.133" cc="" mt=""/>
  <pitch des="Foul" des_es="Foul" id="555" type="S" tfs="031238" tfs_zulu="2016-11-03T03:12:38Z" x="73.55" y="169.15" event_num="555" on_2b="446386" sv_id="161102_231547" play_guid="0ad971b5-f3e6-48a7-864e-
  5d4105645882" start_speed="99.6" end_speed="92.6" sz_top="3.3" sz_bot="1.48" pfx_x="5.48" pfx_z="10.04" px="1.14" pz="2.579" x0="1.08" y0="50.0" z0="6.123" vx0="-1.891" vy0="-145.881" vz0="-8.484"
  ax="11.947" ay="30.852" az="-10.224" break_y="23.8" break_angle="-43.9" break_length="3.1" pitch_type="FF" type_confidence=".896" zone="12" nasty="51" spin_dir="151.443" spin_rate="2485.314" cc="" mt=""/>
  <pitch des="In play, run(s)" des_es="En juego, carrera(s)" id="556" type="X" tfs="031324" tfs_zulu="2016-11-03T03:13:24Z" x="128.93" y="195.18" event_num="556" on_2b="446386" sv_id="161102_231628"
  play_guid="013f9eac-3005-4a7c-af8b-29c5f6678c63" start_speed="97.1" end_speed="90.7" sz_top="3.3" sz_bot="1.48" pfx_x="4.99" pfx_z="9.51" px="-0.313" pz="1.615" x0="0.909" y0="50.0" z0="6.023"
  vx0="-5.281" vy0="-142.002" vz0="-10.245" ax="10.341" ay="27.709" az="-12.383" break_y="23.9" break_angle="-33.1" break_length="3.3" pitch_type="FF" type_confidence=".826" zone="7" nasty="44"
  spin_dir="152.414" spin_rate="2281.354" cc="" mt=""/>
  <runner id="446386" start="2B" end="" event="Home Run" event_num="561" score="T" rbi="T" earned="T"/>
  <runner id="434658" start="" end="" event="Home Run" event_num="561" score="T" rbi="T" earned="T"/>
</atbat>
▼<atbat num="68" b="1" s="0" o="2" start_tfs="031418" start_tfs_zulu="2016-11-03T03:14:18Z" batter="424825" stand="R" b_height="5-10" pitcher="547973" p_throws="L" des="Coco Crisp singles on a ground ball to
left fielder Ben Zobrist. " des_es="Coco Crisp pega sencillo con rodado a jardinero izquierdo Ben Zobrist. " event_num="566" event="Single" event_es="Sencillo" play_guid="b3f55581-6f58-4b12-b9fd-
3a1b6d597773" home_team_runs="6" away_team_runs="6">
```


Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE

Fielding Metrics

- First step: Measure the time elapsed from time of bat-on-ball contact to the fielder's first movement toward the ball.
- First step efficiency: Measures the angle of deviation from a straight line to the ending point of a batted ball trajectory vs. the actual initial path taken toward the ball.
- Max speed: Measures the maximum speed at any point while tracking any ball hit into play.
- Acceleration (outfield): Measures the time elapsed from time of bat-on-ball contact to max speed at any point while pursuing any ball hit into the outfield.
- Total distance: The total distance covered from batted ball contact to fielding the ball.
- Arm strength: Measures the maximum velocity of any throw made by any fielder.
- Exchange: Measures the time from the point a fielder receives the ball to releasing a throw.
- Pop time: Measures the time elapsed from a pitch reaching catcher's glove, to throw, to receipt of the ball by fielder at the intended base on all pickoff throws and steal attempts.
- Pivot: Measures the time elapsed between receipt of the ball and release of throw on double-play attempts.
- Route efficiency (outfield): Divide the distance covered by the fielder by a straight-line distance between the player's position at batted ball contact and where the ball was fielded.

Baserunning Metrics

- Lead distance: Measures the distance between the base and the runner's center of mass at the time the pitcher goes into his windup on a pitch or pickoff attempt.
- Secondary lead: Measures the distance between the base and the runner's center of mass when the ball is released by the pitcher on a pitch or pickoff attempt.
- First step: Measures the time elapsed from time of bat-on-ball contact to the runner's first movement toward next base.
- Stealing first step: Measures the time elapsed from the pitcher's first movement in the stretch to the runner's first movement toward the next base on a steal attempt.
- Acceleration: Measures the time elapsed from time of bat-on-ball contact to the runner's max speed at any point ball is in play.
- Max speed: Measures the maximum speed at any point for all players while the ball is in play.
- Dig speed: Measures the time from bat-on-ball contact to the point where the batter-as-runner reaches first base on an infield ground ball.
- Extra bases: Measures the time of bat-on-ball contact to the point the runner advances an "extra" base (first to third or home, or second to home) on all hits (excluding over-the-fence home runs).

Hitting Metrics

- Exit velocity: Velocity of the ball off the bat on batted balls.
- Launch angle: The vertical angle at which the ball leaves the bat on a batted ball.
- Vector: Classifies the horizontal launch direction of the batted ball into five equal zones of 18 degrees each.
- Hang time: Measures the time from bat contact to the ball either hitting the ground/wall or contact by a fielder.
- Hit distance: Calculates the distance on the ground of the actual landing point of any ball hit into play, ground/wall or contact with fielder, regardless of outcome.
- Projected HR distance: Calculates the distance of projected landing point at ground level on over-the-fence home runs.

Pitching Metrics

- Release: Measures the time from pitcher's first movement out of the stretch to the release point of the pitch.
- Extension: Measures the distance of the release point of the pitch from the front edge of the pitching rubber.
- Velocity: Measures the peak velocity of a pitch at any point from its release to the front edge of home plate.
- Perceived velocity: Velocity of the pitch at the release point normalized to the average release point for MLB pitchers. For example, a 90-mph pitch at a 54-inch release point will seem faster to the batter than a pitch of the same velocity thrown from a 56-inch release point.
- Spin rate: Measures the spin rate of the ball at the point of the release from the pitcher's hand.

Raving NEXT
INDIAN GAMING ANALYTICS AND MARKETING CONFERENCE