

Una introducción a la analítica deportiva

Esteban Navarro Garaiz

estebanng.github.io



@EstebanNG_

Esteban



UNAM

2013 - 2018

Licenciatura en Matemáticas



New York University

2018 - 2020

Maestría en Ciencia de Datos



Los Angeles Dodgers

2020 - 2021

Analista Cuantitativo



TEAMWORKS

Teamworks (Zelus Analytics)

2022 - Presente

Gerente Sénior de Producto Técnico





Óscar Hurtado, el físico mexicano y su aportación en el título de Rangers

Zelus Analytics

70+ plantilla

20+ PhDs

30+ equipos

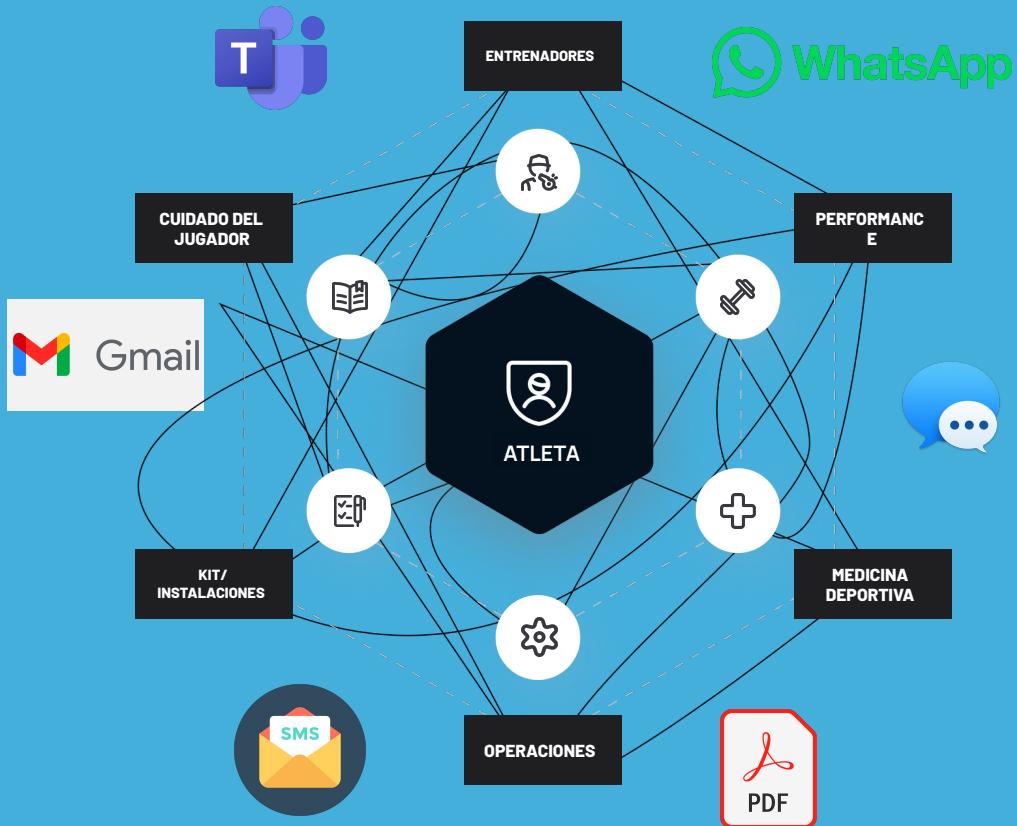


Equipos donde trabajaron empleados de Zelus:



El Desafío

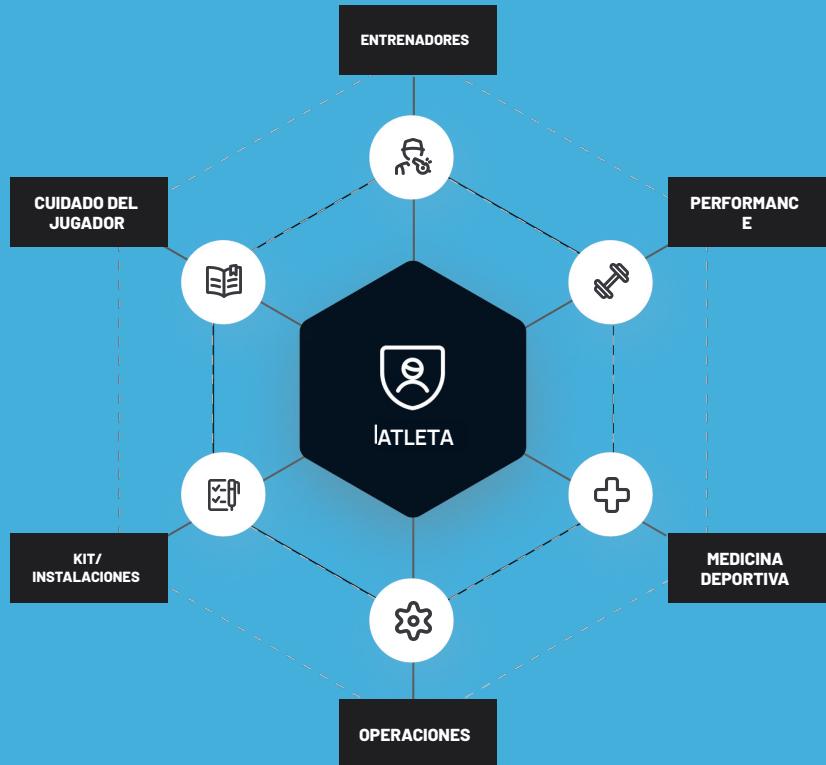
En la medida que los departamentos deportivos crecen en complejidad y tamaño, surgen **problemas de comunicación**. Los departamentos funcionan solos, con procesos y sistemas desconectados.





TEAMWORKS

El sistema operativo para deportes



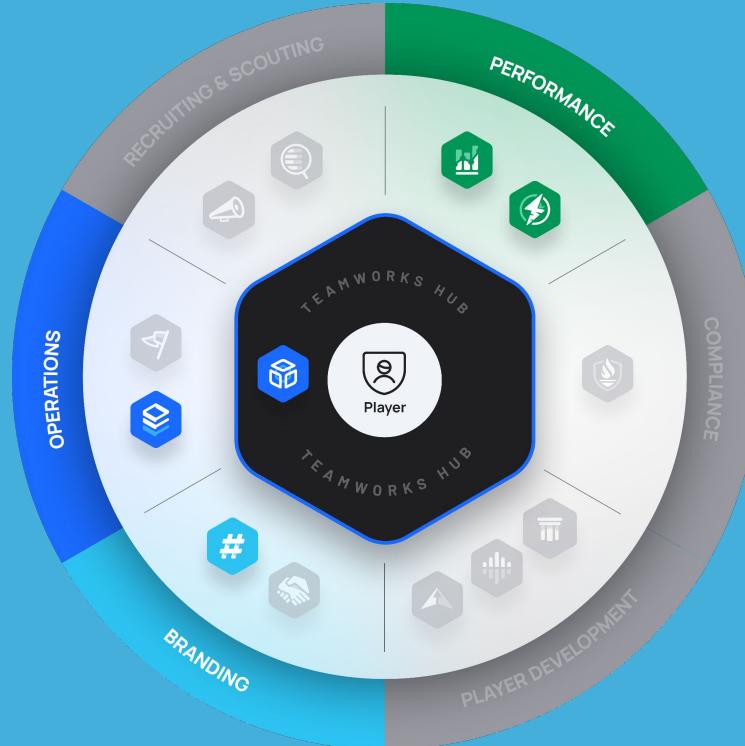


TEAMWORKS

El sistema operativo para deportes

Una plataforma simple, consolidada e integrada diseñada específicamente para organizaciones deportivas élite.

- **Hub:** Comunicación, planificación y logística en un sólo lugar
- **AMS:** Plataforma para optimización y monitoreo de atletas



1.

¿Qué es la analítica deportiva?

¿Y para qué se usa?

Caso de estudio: Pete Alonso

Edad (abril 2025)

30.3

~30%
mejor
promedio

wRC+

131

WAR ('21-'24)

12.2

7mo 1B
MLB

Posición

1B

Defensa

👎 👎

Fans

😍 😍

Proyecciones de contratos:

- MLBTR: 5 años / \$125M
- Jim Bowden: 7 años / \$189M



Caso de estudio: Pete Alonso



Proyecciones de contratos:

- **MLBTR: 5 años / \$125M**
- **Jim Bowden: 7 años / \$189M**

Consideraciones:

- **Presupuesto**
- **Proyección a futuro**
- **Compatibilidad en roster**
- **Costo de oportunidad**
 - **Monetario**
 - **Espacio en roster**
- **¿Es mi equipo suficientemente bueno?**
- **¿Es un buen contrato?**

Desafíos para operar un equipo de béisbol



Construcción de roster

1. Evaluación de talento
 - a. FA / Draft / cambios
2. Optimización de recursos
3. Interpretabilidad



Planeación de partidos

1. Alineación y cambios
2. Entendimiento oponente
3. Estrategia de lanzadores
4. Posicionamiento defensivo



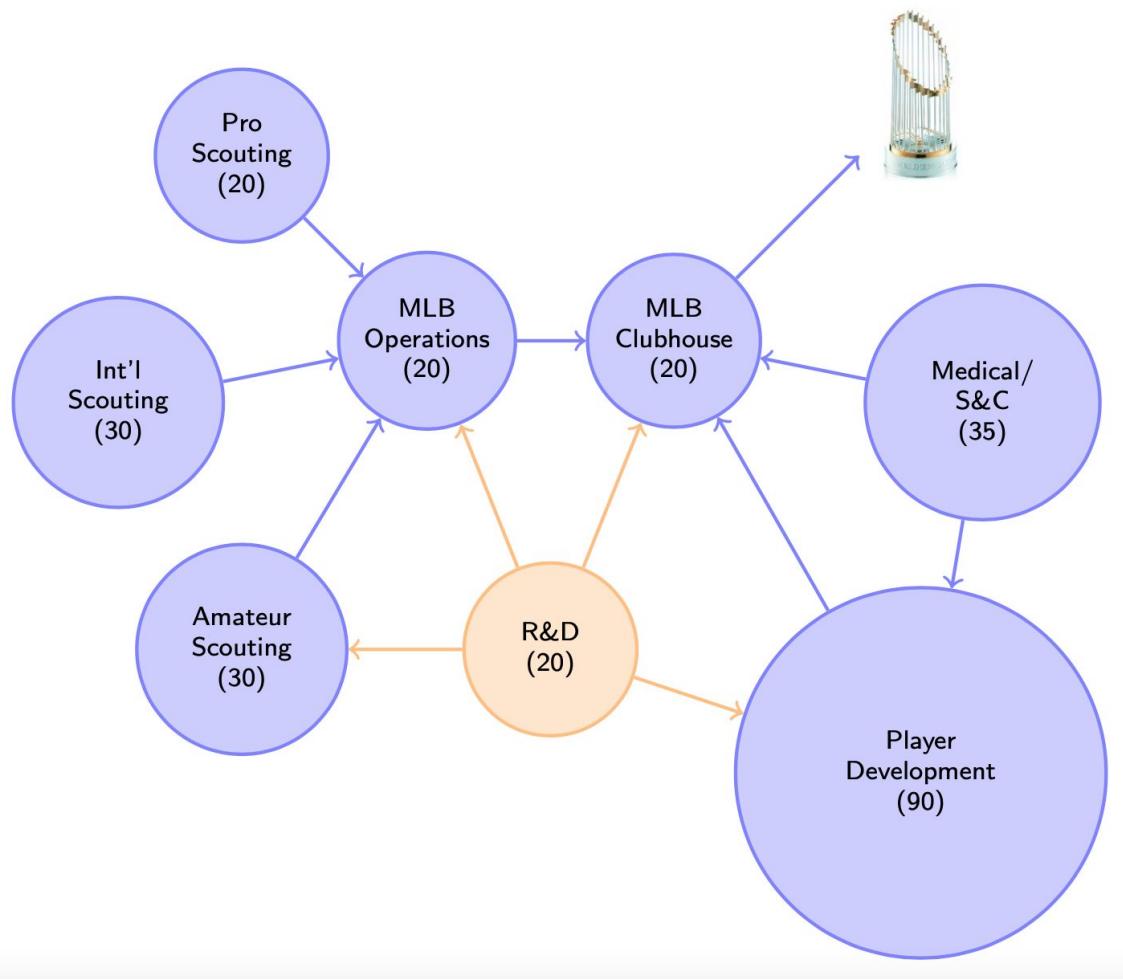
Desarrollo de jugadores

1. Prevención lesiones
2. Desarrollo personalizado
3. Crear y monitorear métricas
4. Promoción filiales



Operaciones del equipo

1. Obtención, almacenamiento y transformación de datos
2. Nutrición y gestión médica
3. Logística de viajes y calendarización de eventos





Analítica deportiva 101

Herramientas basadas en datos para tomar decisiones y obtener una ventaja competitiva

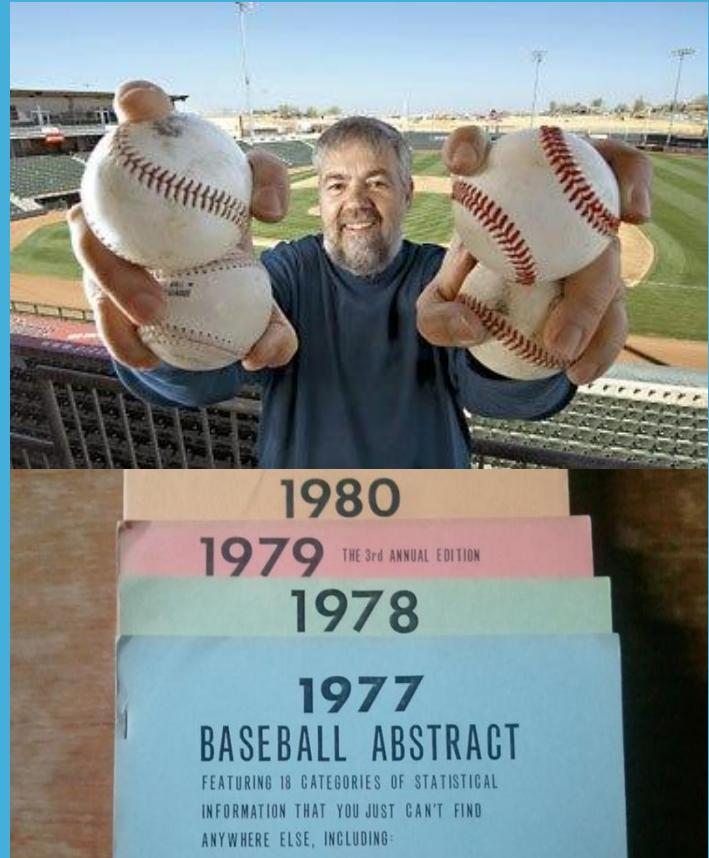
2.

Una breve perspectiva histórica

De Bill James a radares

Bill James

- Economía. PhD en psicología.
- Baseball Abstract (1977)
 - 68 páginas, estudio de estadísticas.
 - Primera edición vendió 75
- En los primeros 5 años define conceptos usados ahora:
 - Curvas de envejecimiento
 - Valor de reemplazo
 - Carreras creadas
 - Equivalencias de ligas mayores



Bill James en Sports Illustrated (1981)

“Ningún jugador en la historia del béisbol ha sido 20 juegos mejor que el jugador promedio en su posición, ni Babe Ruth, ni Ty Cobb, ni nadie.

(...)

La diferencia entre el Jugador Más Valioso y un jugador promedio en su posición es usualmente entre 5 y 8 juegos”

All right, it's quiz time, Baseball 101. True or false:

1) One big reason why Billy Martin has had such remarkable managerial success in reviving listless teams in Minnesota, Detroit, Texas, New York and now Oakland is his genius for inspiring young ballplayers.

2) Another reason for Martin's repeated success is his insistence on having a team that steals bases.

3) O.K., Nolan Ryan loses nearly as many games as he wins, but give him a lead in the late innings and he is incomparable, almost unbeatable.

4) And even though Ryan is paid an enormous salary, he more than earns it as a big gate attraction.

5) Montreal had better watch out for September. The Expos blew their chances of winning the NL East the last two seasons by cooling off in that month.

6) Any analyst who picked the champion Phillies to finish fifth last year ought to turn in his statistics.

There you have it. Easy quiz, isn't it? Every real fan within earshot of Joe Garagiola knows that all these statements are true, right?

Sorry. They're all false:

1) Of Billy Martin's teams, only Oakland and Texas could be called young. By the time of the All-Star break, his first Minnesota squad had only one non-pitching starter under the age of 27. His first Detroit team had only one under 28. His first Yankee team had two under 27, but one, age 26, was replaced by a 31-year-old the next year. The secret must lie elsewhere.

2) Speed! The Twins were third in steals in the American League the season before Martin took them over, but fell to fourth with him. His division champion Tigers stole only 17 bases all season long. The Rangers barely improved their league ranking under Martin, moving from sixth to fifth. The Yankees did jump from sixth to third in steals during Martin's first year as their manager, but fell to seventh the next

continued

HE DOES IT BY THE NUMBERS

The esoteric equation on the Royals' scoreboard in Kansas City is only one of the far-out findings of deep-thinking baseball statistician Bill James

by DANIEL OKRENT



Keith Woolner (2000) – Problemas de Hilbert en el béisbol

1. La analítica es útil en tanto informa una decisión
2. El análisis de béisbol debe considerar el contexto completo del juego
3. Los recursos para responder una pregunta pueden no estar disponibles
4. Un razonamiento es válido aunque aún no pueda cuantificarse con precisión.





Evaluación de pitchers

1. Separar a la defensa en pitcheo y fildeo

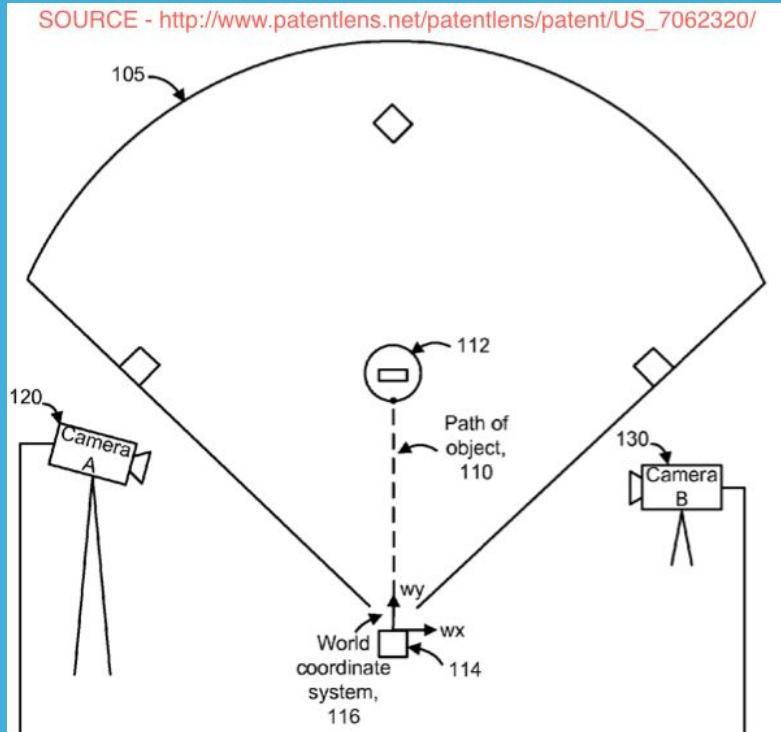
- "El pitcheo y la defensa están tan entrelazados que parecen imposibles de separar."

10. Proyectar con precisión a lanzadores de ligas menores

- "Aunque las [estadísticas] pueden darnos guías aproximadas sobre lanzadores jóvenes buenos y malos, estamos lejos del nivel de certeza que queremos alcanzar."

PitchFX (2006)

- Sistema 2 cámaras
- 20 fotos del lanzamiento
 - Ajusta una trayectoria basado en la ubicación 3D de la pelota en las fotos
- Devuelve características como la velocidad, ubicación, movimiento
- Manualmente, un operador añade el resultado del lanzamiento, el turno al bate, cambios, etc

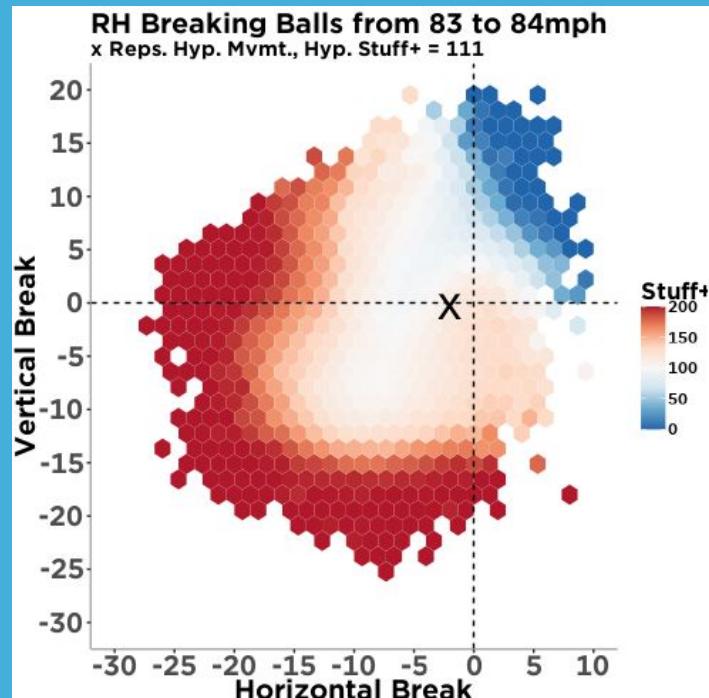


Evaluación de pitchers

Defense-Independent Pitching

- DIPS, creada por Voros McCracken (2001)
- Sólo usa lo que el pitcher controla (K, BB, HR, HBP)
- Popularizada por Tom Tango como FIP (Fielding Independent Pitching) y variantes asociadas

Stuff+



Pitch Design: What is Stuff+?



■ 16. Crear un sistema para evaluar cambios

- Difícil: Hay muchos posibles motivos para hacer un cambio
- Debe evaluarse al momento del cambio
- Considerar:
 1. Factores financieros
 2. Rendimiento de jugadores
 3. Competitividad del equipo / prioridad de la directiva por ganar en ese momento

16. Crear un sistema para evaluar cambios

Valor de mercado

- Valor proyectado por la duración del contrato, traducido a dólares
- Substráe el costo del contrato (proyectado o garantizado)
- Valor total para el equipo

\$36.5M > \$33.8M

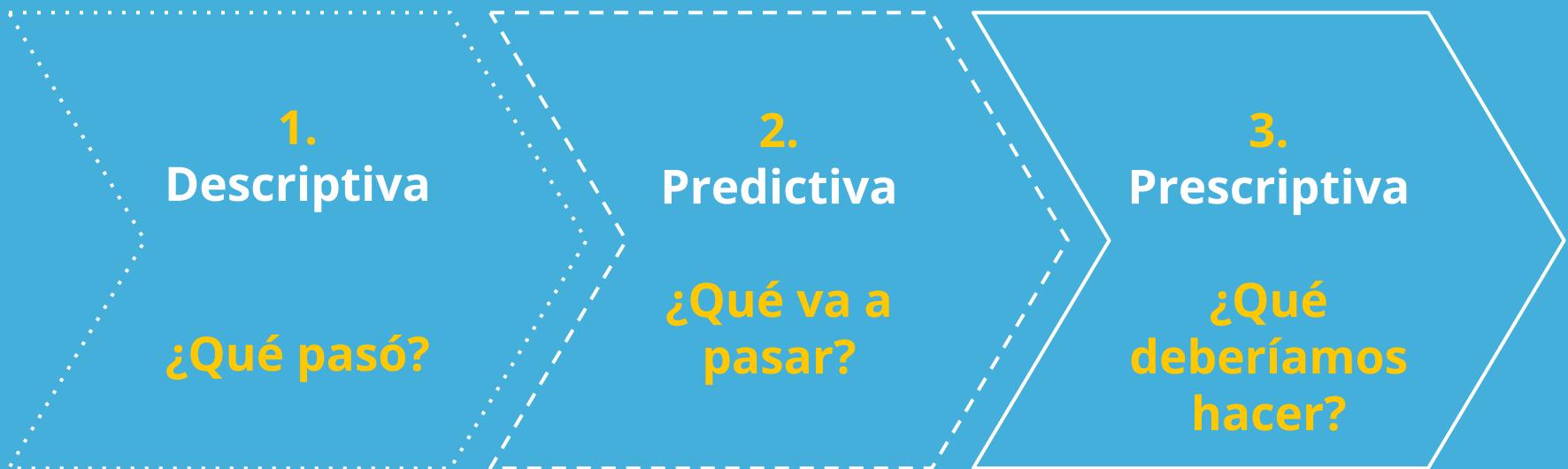


3.

Analítica Deportiva 102

O cómo ganar la Serie Mundial en 3 sencillos pasos

Analítica Deportiva 102



Analítica Deportiva 102

1. Descriptiva

¿Qué pasó?

- Asigna probabilidades a eventos pasados usando patrones
- Estadísticas para explicar el pasado
- ¿Qué medimos? ¿En qué escala?

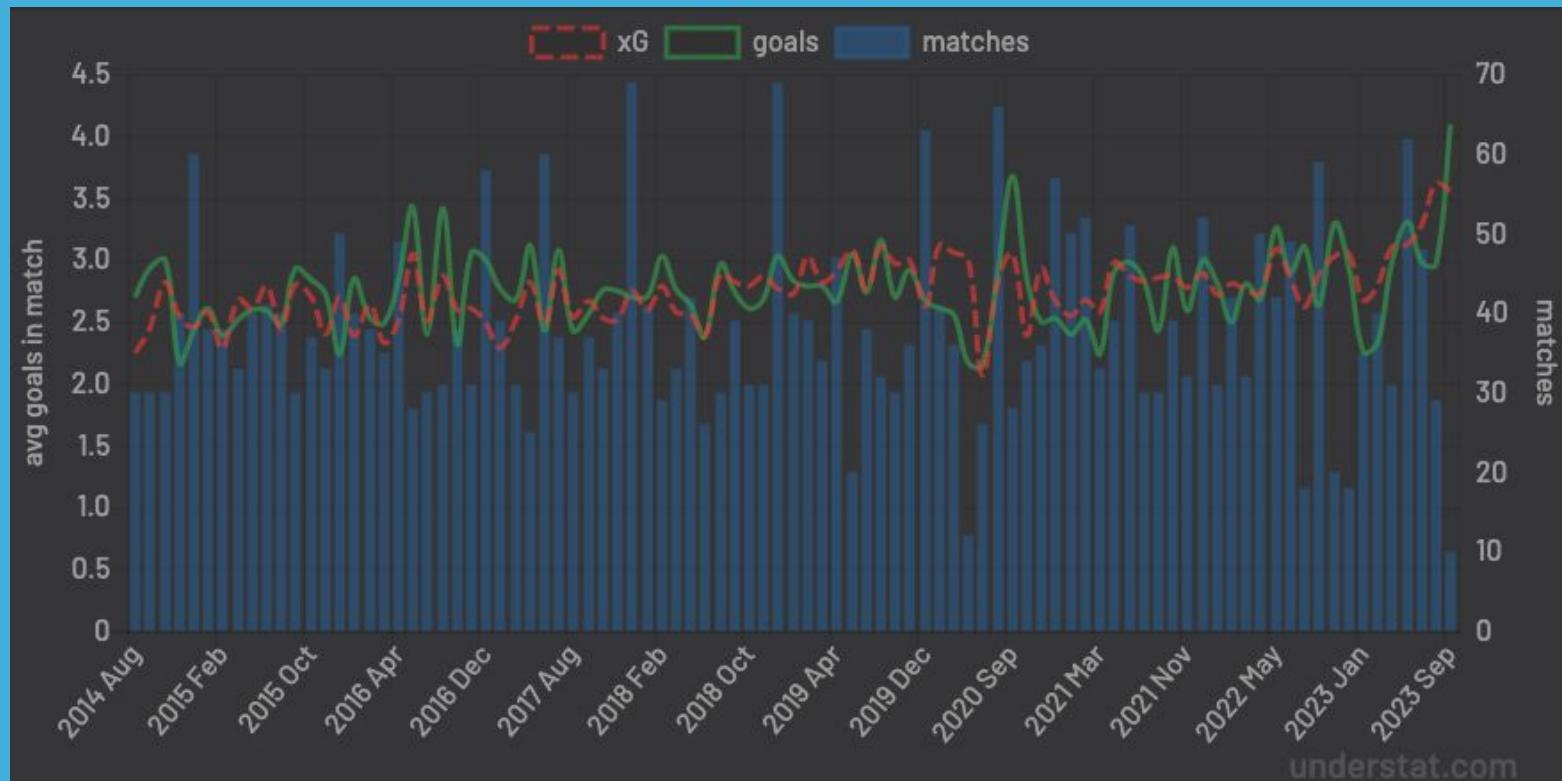
Expected Goals (xG)

- Mide la probabilidad de que un tiro termine en gol
- Datos del tiro. Por ejemplo:
 - Distancia y ángulo a la portería
 - Posición defensas
 - Parte del cuerpo usada
 - Tipo de asistencia (e.g. pase filtrado)



Statsbomb – What Are Expected Goals (xG)?

Expected Goals (xG)



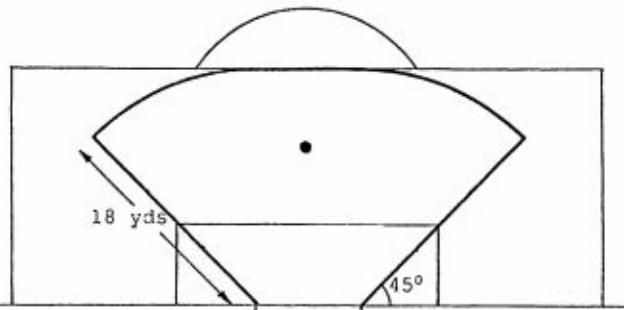


Fig. 2. Location of the arc of relatively high goal scoring probability

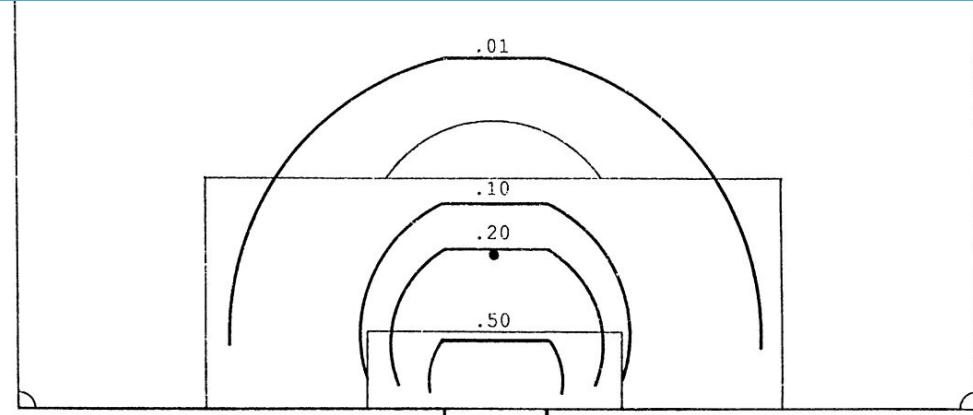
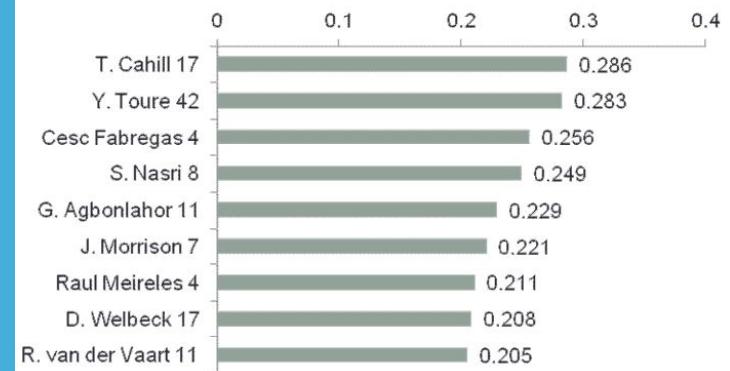


Fig. 3. Scoring probability contours for kicked shots from open play and from less than 1 yard from the nearest opponent



Analítica Deportiva 102

2.
Predictiva
¿Qué va a
pasar?

- Asigna probabilidades a eventos futuros
- Separa suerte de habilidad en residuales descriptivos y atribuye valor a jugadores
- Crea proyecciones para jugadores



Butler P: 15
4. Trumbo 1 FOR 3
BAL 5 
TEX 13 
▲ 6 2 Outs 2-2

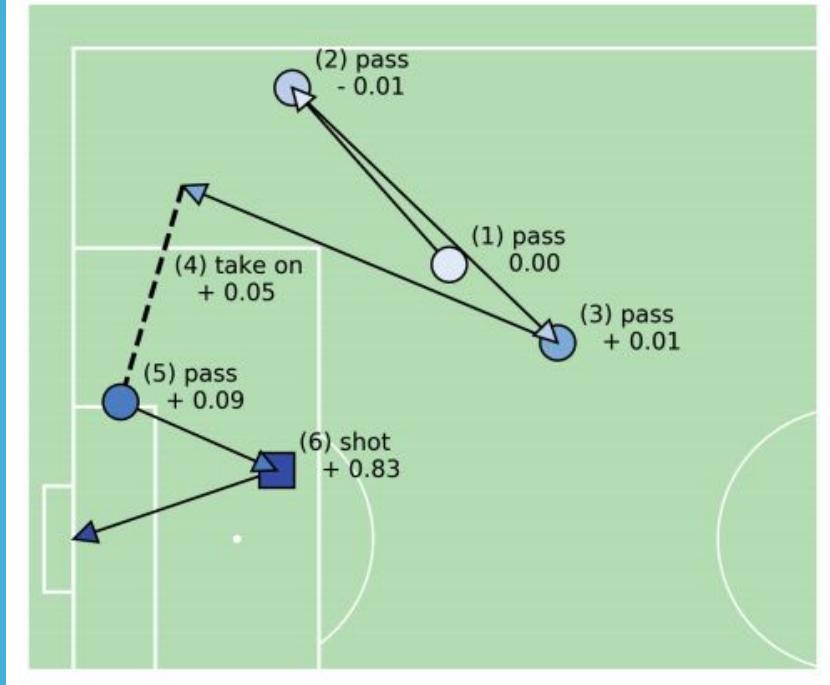
90% outs sobre promedio

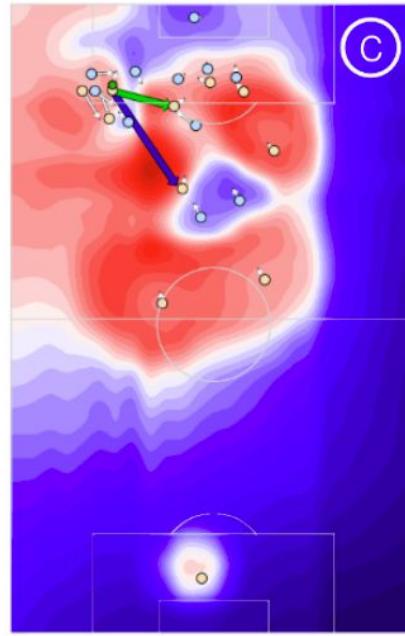
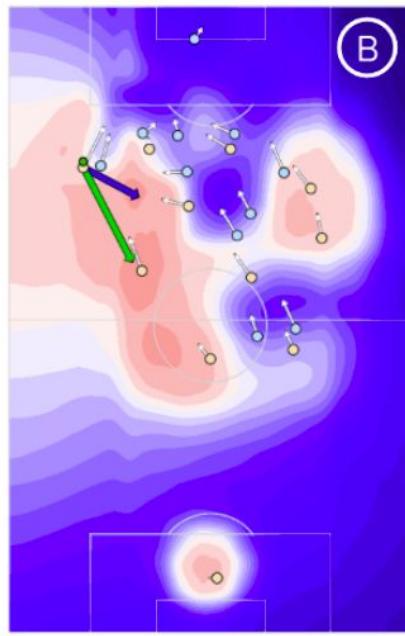
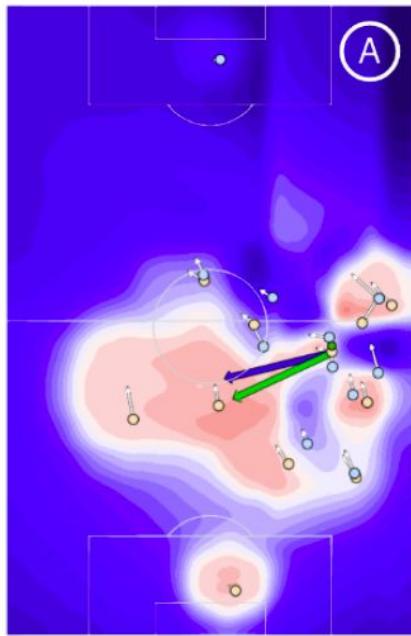


Player	Team	Attempts	OAA
Robert Jr., Luis	S	362	12 
Doyle, Brenton	C	334	12 
Rodríguez, Julio	S	330	11 
Kiermaier, Kevin	B	237	10 

[Baseball Savant -- Outfield Directional OAA Leaderboard](#)

	TIME	PLAYER	ACTION	P_{scores}	VALUE
1	92m4s	S. Busquets	pass	0.03	0.00
2	92m6s	L. Messi	pass	0.02	- 0.01
3	92m8s	S. Busquets	pass	0.03	+ 0.01
4	92m11s	L. Messi	take on	0.08	+ 0.05
5	92m12s	L. Messi	pass	0.17	+ 0.09
6	92m14s	A. Vidal	shot	1.00	+ 0.83



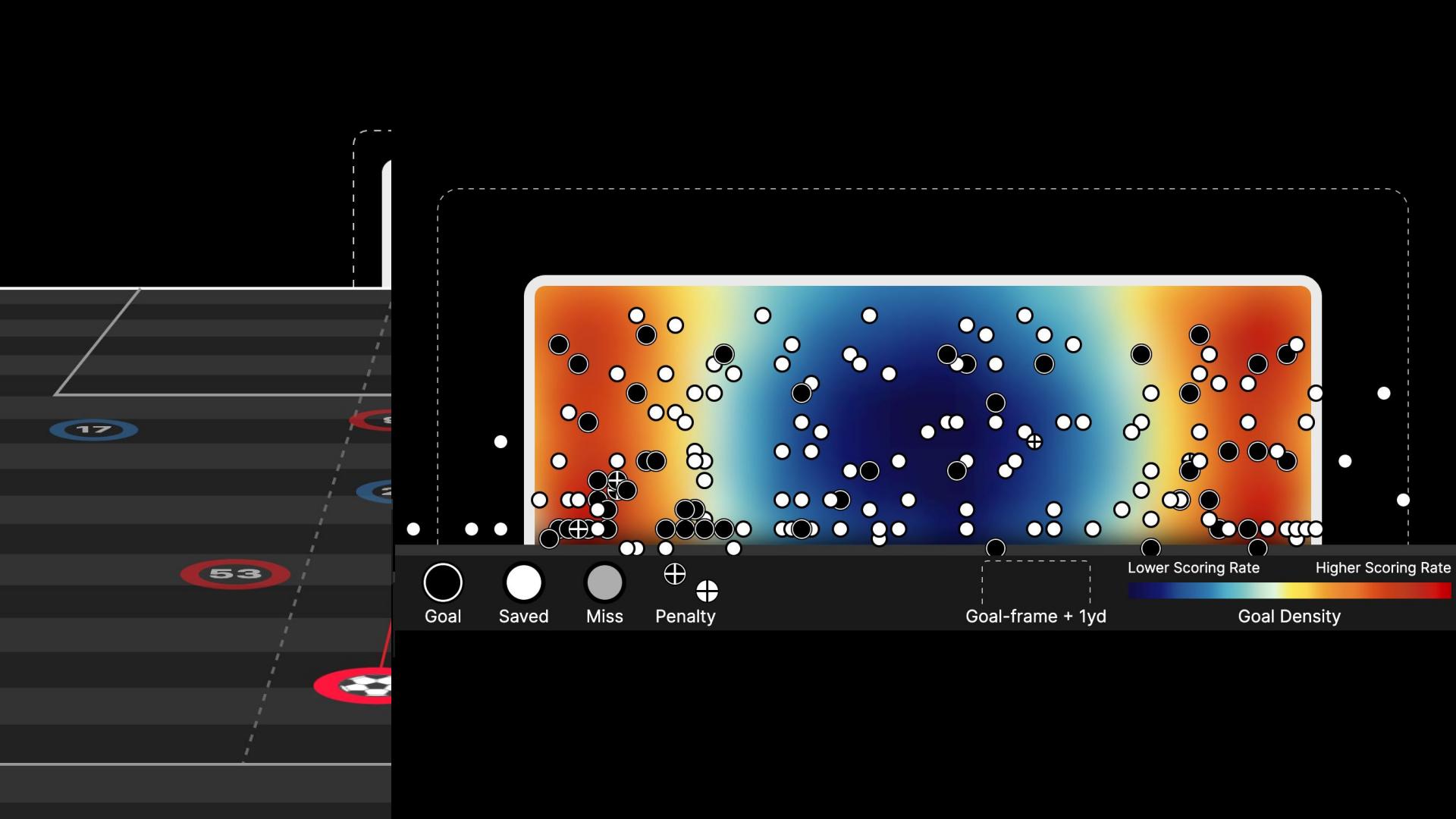


Backward pass releases high pressure, increasing value
(0.044)

Long forward drive maintains EPV
until it increases with
backward pass to Messi
(0.034)

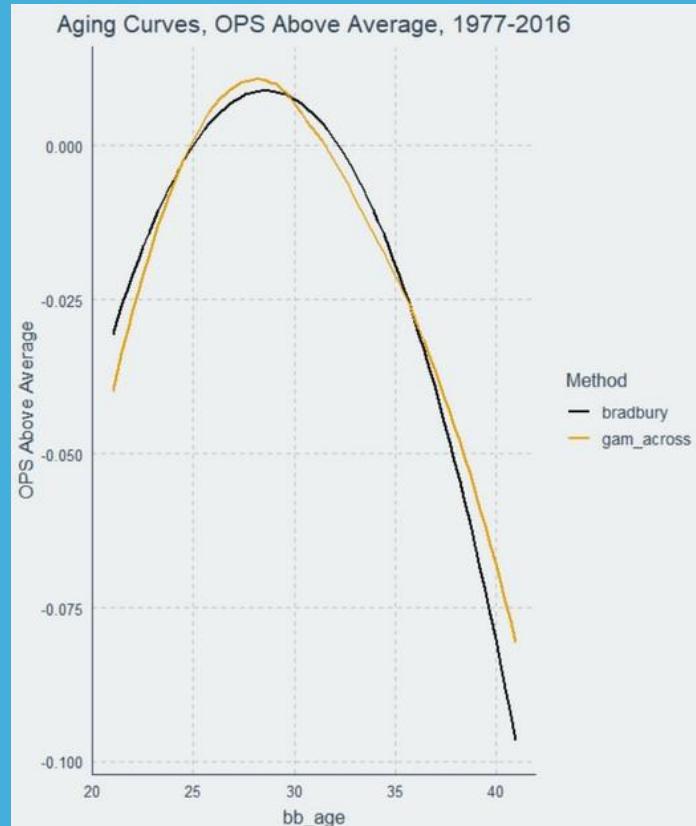
Two short passes inside the box
increases available space. EPV increases shortly and
then decreases rapidly due to high pressure.
(0.085)





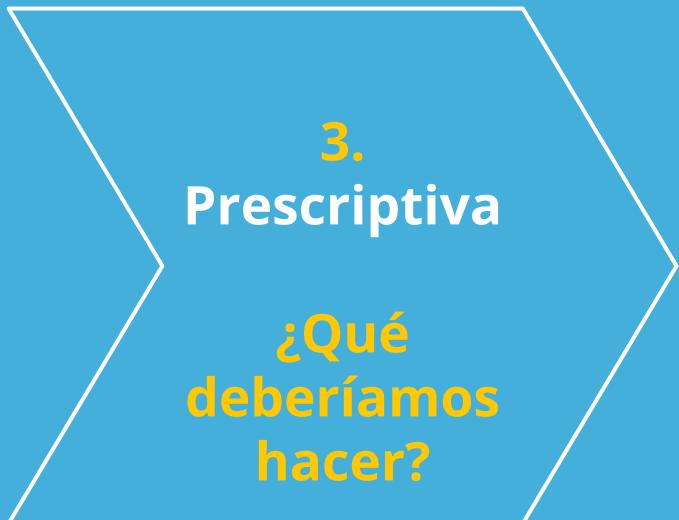
Proyecciones

- ¿Qué de lo observado se va a repetir en el futuro? ¿Cómo va a cambiar el atleta?
- Remover el contexto Contexto
 - Dificultad oponentes
 - Liga dónde se juega  → 
- Regresión a la media
- Curvas de envejecimiento



[Baseball Prospectus – Rethinking Aging Curves](#)

Analítica Deportiva 102



- Asigna acciones a proyecciones
- Se usa para toda clase de toma de decisiones bajo incertidumbre
- Crea información accionable al contextualizar modelos

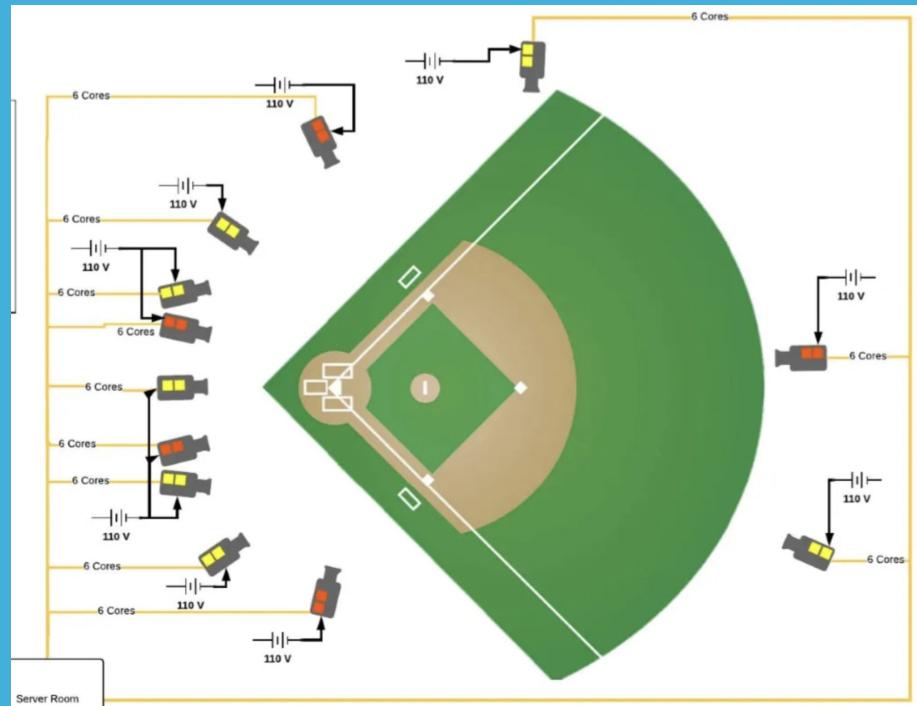
4.

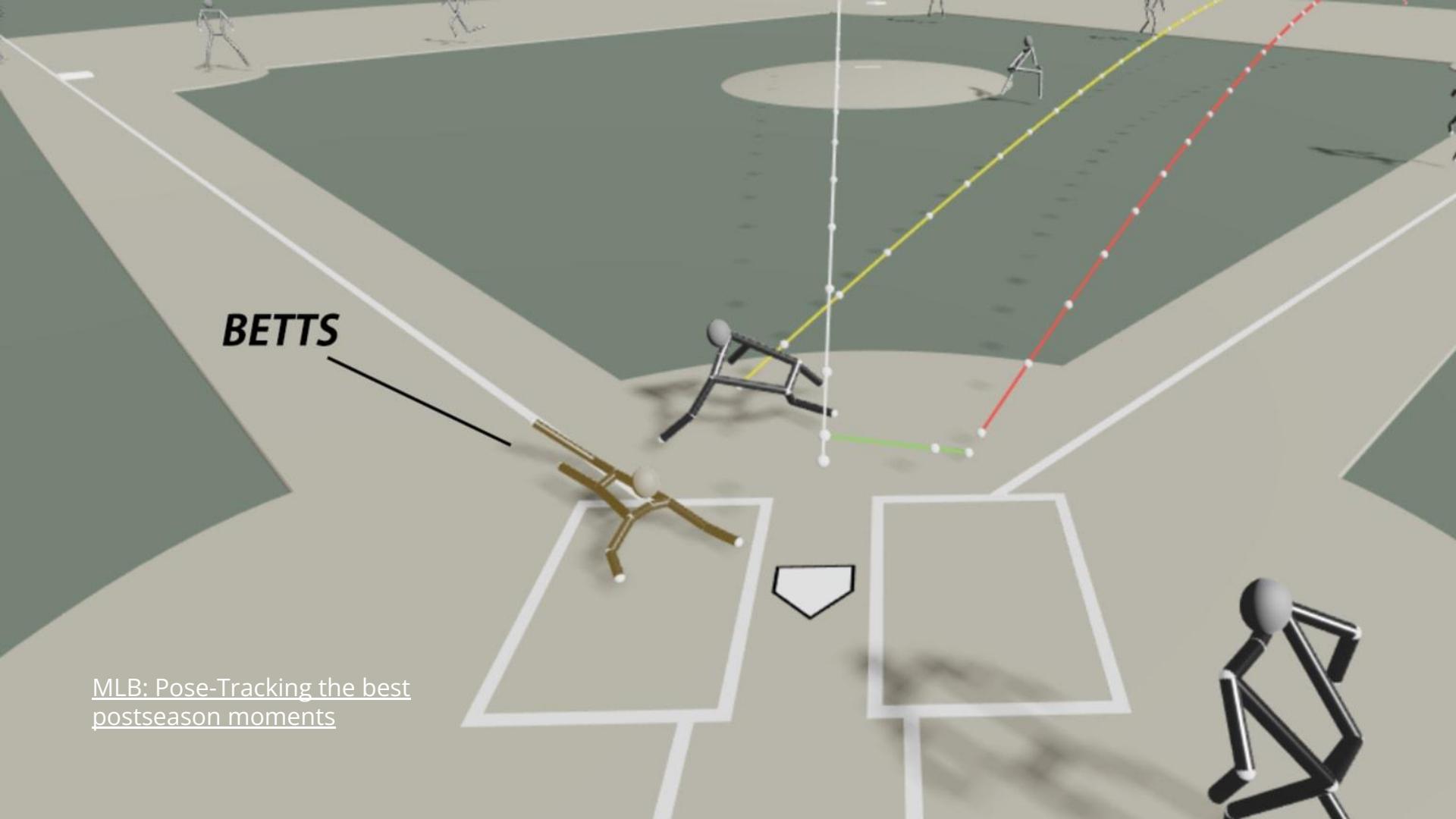
El futuro de la analítica deportiva

A través de dos problemas en los que estoy emocionado en trabajar

Hawk-Eye (2020)

- Sistema de 12 cámaras ópticas
- ¡Muchísimos más datos!
- Poses esqueléticas de 29 puntos a 300 cuadros por segundo

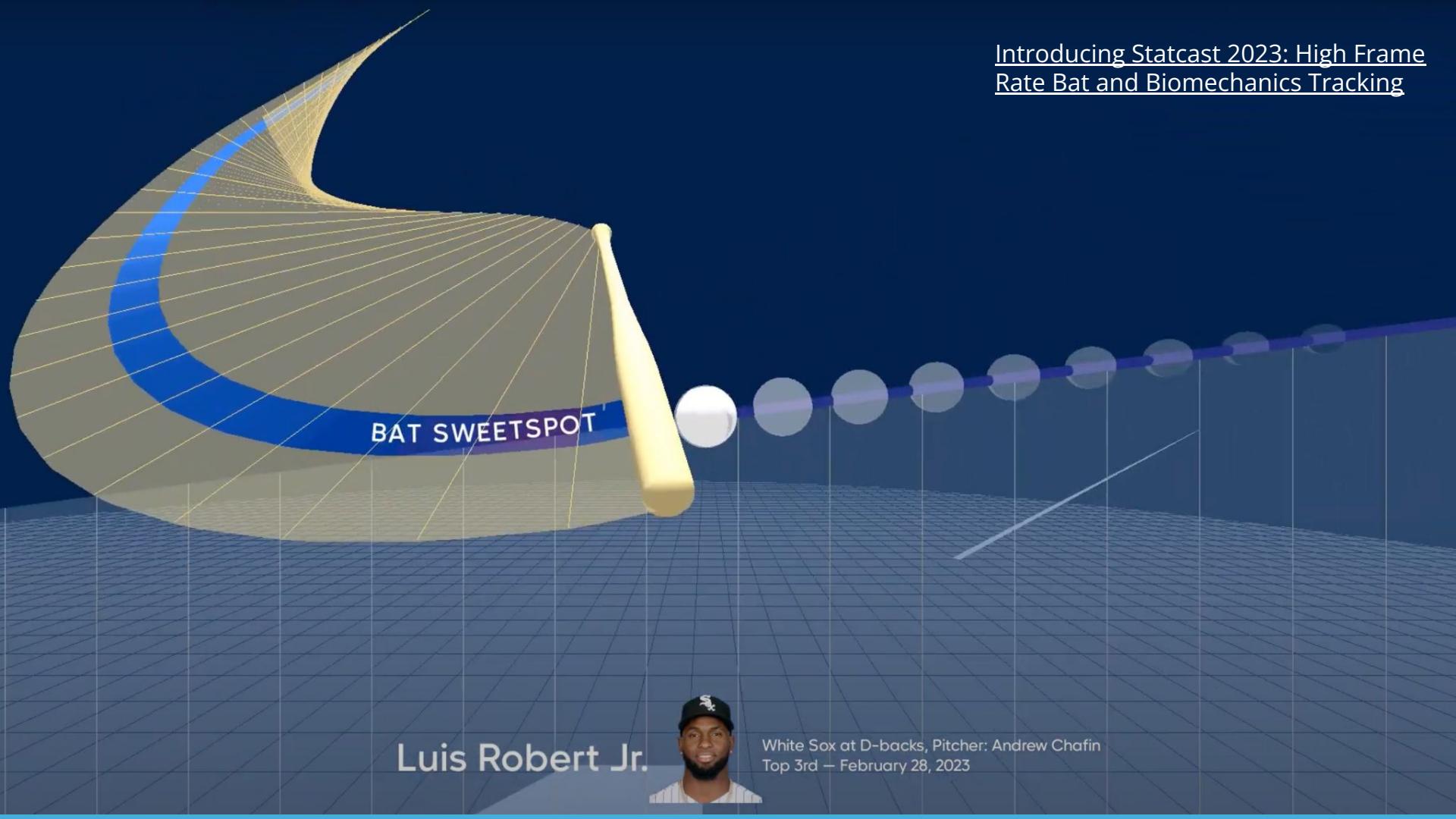




BETTS

MLB: Pose-Tracking the best
postseason moments

Introducing Statcast 2023: High Frame
Rate Bat and Biomechanics Tracking



Luis Robert Jr.



White Sox at D-backs, Pitcher: Andrew Chafin
Top 3rd – February 28, 2023

¿Qué es un buen swing?

- Mecánica de bateo de Aaron Judge (2.01m) v.s. José Altuve (1.68m)
- Proyecciones de bateadores en ligas menores
 - Batear = reaccionar al lanzamiento
 - Dificultad enfrentada
- Explotación de debilidades (ej. rectas altas)
- Reacción pura v.s. adivinar con información
- Varianza en movimientos físicos



Preparación y desarrollo de jugadores con datos

- Equilibrio entre:
 - Corto y largo plazo
 - Cambios/mejoras y riesgo de lesión
 - Carga de trabajo y propensión a lesiones
- Cuándo intervenir como entrenador
- En béisbol
 - Disponibilidad de pitchers, dado su carga en los últimos juegos
 - Entender mejor el efecto de enfrentar a un pitcher 2+ veces



Recursos Técnicos

Una empujadita para conseguir el trabajo de tus sueños

¿Cómo me contratan?

Programar



Tecnología Nube



Mantenimiento Modelos



Comunicación

- Audiencia técnica y no técnica
- Entender el problema

Modelos Estadísticos

- Modelos lineales generalizados
- Machine Learning

Proyectos Públicos

A screenshot of a GitHub profile page for 'edwebster / edwmc_v1'. The page shows a profile picture of a man, a bio about his work in data science, machine learning, statistics, data engineering, and data visualization, and links to his LinkedIn and GitHub profiles. It also lists his repositories, including 'fotball_analytics'.

¡Recursos para Analítica del Deporte!

Soccer

- [Devin Pleuler's Soccer Analytics Handbook](#)
- [Devin Pleuler's NorthPitch](#)
- [Statsbomb](#)
- [Metrica Sports](#)
- [Karun Singh's Expected Threat](#)

Béisbol

- [Pybaseball / BaseballR](#)
- [RetroSheet](#)
- [Lahman Database](#)

Conferencias

- [Women in Sports Data](#)
- [New England Symposium on Statistics in Sports \(NESSIS\)](#)
- [Carnegie Mellon Sports Analytics Conference](#)
- [Uconn Sports Analytics Symposium](#)
- [SaberSeminar](#) 
- **Misceláneos**
- [Kaggle Data Sets](#)
- [Sports Viz Sunday](#)

Futbol Americano

- [NFL FastR](#)
- [NFL ScrapR](#)
- [Pro Football Reference](#)

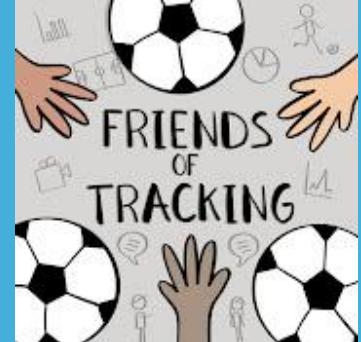
Tenis

- [Stats On the T](#)

Básquetbol

- [Basketball Reference](#)
- [HoopR / Py_ball](#)
- [Wehoop](#)

Friends of Tracking

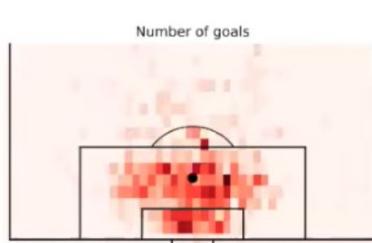


- Canal de Youtube, empezado en la pandemia con:
 - Tutoriales: Mapas de pase, modelo control de campo, etc.
 - Discusiones, conferencias
- Código disponible en Github
- ¡Javier! (y más)





goals/shots = frequency of scoring

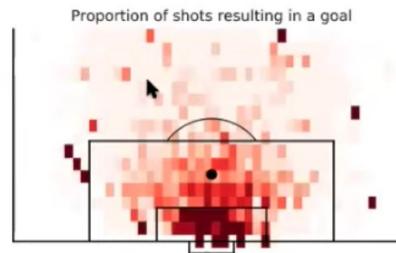


14
12
10
8
6
4
2
0



50
40
30
20
10
0

=

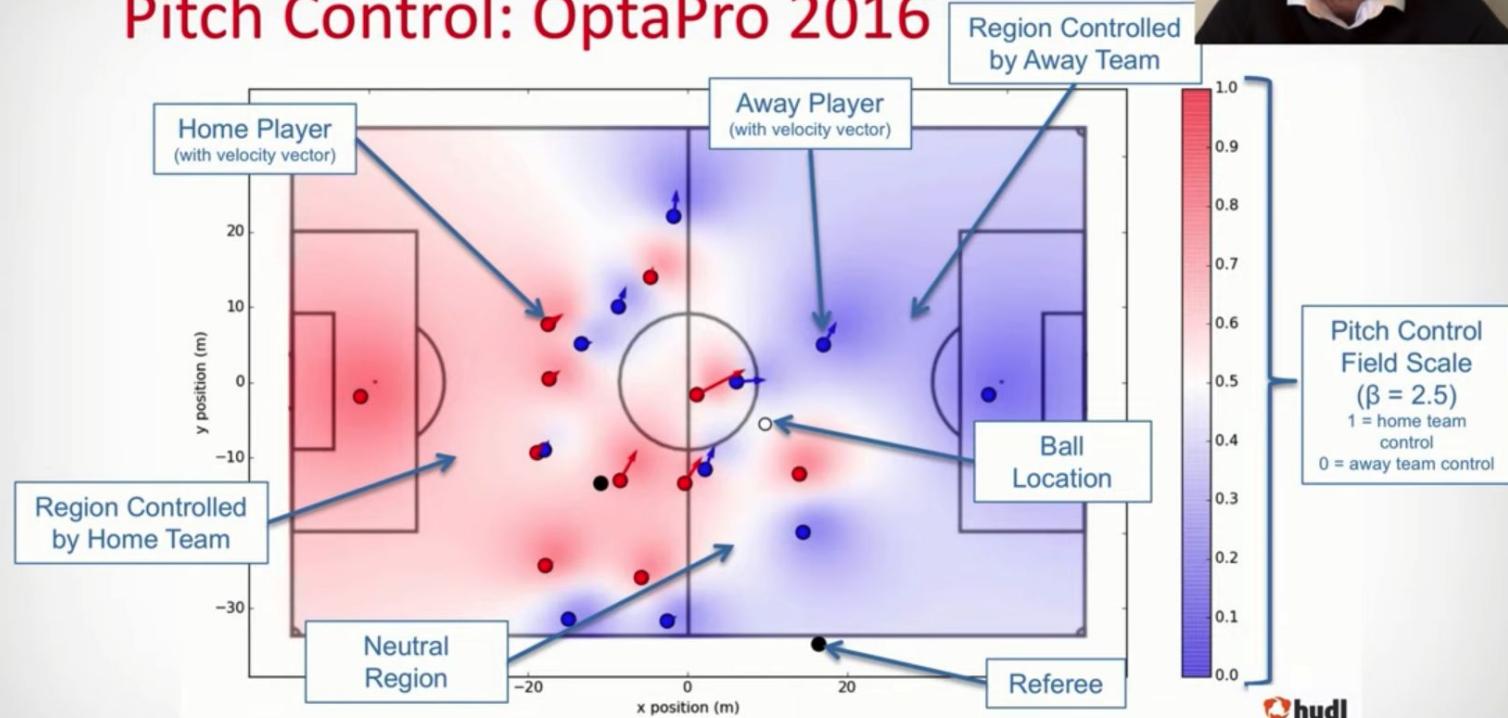


0.5
0.4
0.3
0.2
0.1
0.0





Pitch Control: OptaPro 2016



FoT: Liverpool FC data scientist William Spearman's masterclass in pitch control



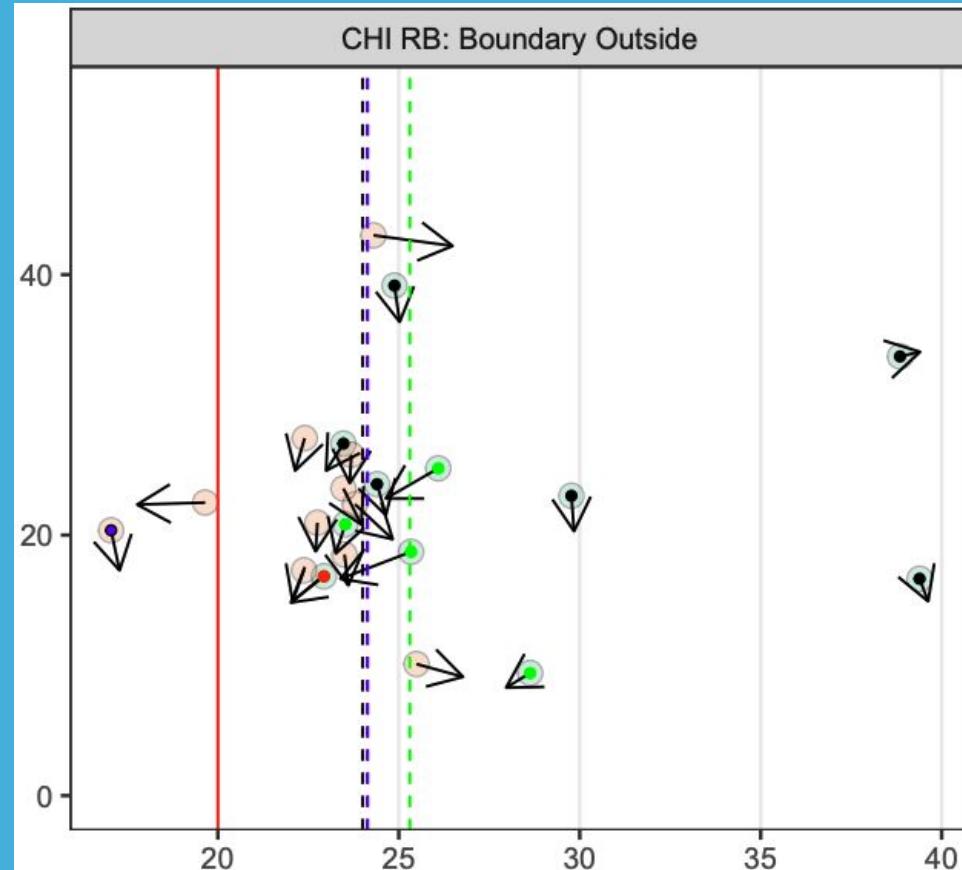
Big Data Bowl

- Hackathon organizado por la NFL desde 2019
- Edición 2024: \$100,000 USD en premios
- ¡Empezó en octubre! 🎉
- Mentor profesional para el equipo

Big Data Bowl (2020)



- Objectivo: predecir número de yardas ganadas por tierra
- Datos: inicio de la jugada
 - Ubicación jugadores
 - Velocidades
 - Aceleraciones



Matt Ploenzke – Big Data Bowl (2020)



PySport

open source

 PySport
open source

Filters ▾ [New projects!](#) [Last changed](#)

Search

Languages	Sports	Category
<input type="checkbox"/> Haskell	<input type="checkbox"/> American Football	<input checked="" type="checkbox"/> Scraper/API
<input checked="" type="checkbox"/> Python	<input type="checkbox"/> Australian Football	<input checked="" type="checkbox"/> Model/Calculations
<input type="checkbox"/> R	<input type="checkbox"/> Baseball	<input checked="" type="checkbox"/> IO (Reading/Writing)
<input type="checkbox"/> Other	<input type="checkbox"/> Basketball	<input checked="" type="checkbox"/> Visualization
	<input type="checkbox"/> Chess	<input checked="" type="checkbox"/> Open-data
	<input type="checkbox"/> Cricket	<input checked="" type="checkbox"/> Database
	<input type="checkbox"/> Cycling	
	<input type="checkbox"/> Field Hockey	
	<input type="checkbox"/> Ice Hockey	
	<input type="checkbox"/> Motorsport	
	<input type="checkbox"/> Netball	
	<input type="checkbox"/> Soccer	
	<input type="checkbox"/> Tennis	

[mit your package here](#)

matches found



¡Muchas gracias!
Esteban Navarro Garaiz

estebanng.github.io

 @EstebanNG_