

# Análisis de Datos Futbolísticos

Javier Bustos

# Tipos de Modelamiento

- Nivel 1: estadísticas básicas (sumas de minutos, goles, tarjetas, citaciones, etc.)
- Nivel 2: a nivel de eventos (sigo a quien lleva la pelota y veo qué hace)
- Nivel 3: análisis de videos/GPS (movimientos del equipo e individuales)

# Datos con los que cuento

- Nivel 1: planillas de partidos, datos de tests físicos estandarizados, datos fisiológicos (nutricionista)
- Nivel 2: TAG de videos -> eventos (se conoce como LOG)
- Nivel 3: TAG automático de participantes + posición + ubicación del balón + ubicación del árbitro + ubicación de peloteros + ubicación de la barra + ubicación de las bancas + ...

# Modelamiento de futbolistas

## Nivel 1

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 1: GPIV

- ¿Quién hace los goles importantes?
- Equipos similares pueden lograr diferencia de +/- 2 goles (más que eso es uno superior al otro => probabilidad bajísima de "darlo vuelta")
- Importancia del gol: aumenta mi probabilidad de ganar (o empatar)
- Después del 3er gol de diferencia ya da lo mismo así que todos esos no suman
- En FF chileno  $P(\text{gana local}) \sim P(\text{gana visita}) \sim 0.45$  antes de jugar

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 1: GPIV

- $P(\text{ganar} \mid \text{está } 3+ \text{ goles arriba}) = 1.0$ ,  $P(\text{perder} \mid \text{está } 3+ \text{ goles arriba}) = 0.0$
- $P(\text{ganar} \mid \text{está } 3+ \text{ goles abajo}) = 0.0$  ,  $P(\text{perder} \mid \text{está } 3+ \text{ goles abajo}) = 1.0$
- Se calcula con todos los partidos para : +2, +1, 0, -1, -2 goles de diferencia
- Para cada jugadora se calcula:

$$\text{GPIV}(x) = 3 \times [P(\text{ganar} \mid \text{dif de gol ahora}) - P(\text{ganar} \mid \text{dif de gol antes})]$$

$$+1 \times [P(\text{empatar} \mid \text{dif de gol ahora}) - P(\text{empatar} \mid \text{dif de gol antes})]$$

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 1: GPIV (2023)



# ¿Qué puedo hacer con los datos que tengo?

## Nivel 1: Score +/-

- Usado desde ~2010 en NBA y ~2012 en NHL
- "*¿Cómo varía el parámetro objetivo (goles) con la jugadora X en cancha comparado con la jugadora X fuera de la cancha?*"

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 1: Score +/-

- Usado desde ~2010 en NBA y ~2012 en NHL
- "*¿Cómo varía el parámetro objetivo (goles) con la jugadora X en cancha comparado con la jugadora X fuera de la cancha?*"
- En cancha: Primer partido X jugó 60 minutos, pierden 0-1. Segundo partido X entró a los 60' cuando ganaban 3-0 y el partido termina 5-0 =>  $((-1/60) + (2/30)) \times 90 = +4.5$  gpp

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 1: Score +/-

- Usado desde ~2010 en NBA y ~2012 en NHL
- "*¿Cómo varía el parámetro objetivo (goles) con la jugadora X en cancha comparado con la jugadora X fuera de la cancha?*"
- En cancha: Primer partido X jugó 60 minutos, pierden 0-1. Segundo partido X entró a los 60' cuando ganaban 3-0 y el partido termina 5-0 =>  $((-1/60) + (2/30)) \times 90 = +4.5$  gpp
- Fuera de cancha: Primer partido X no estuvo 30 minutos, y en el segundo 60 minutos =>  $((0/30) + (3/60)) = +4.5$  gpp

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 1: Score +/-

- Usado desde ~2010 en NBA y ~2012 en NHL
- "*¿Cómo varía el parámetro objetivo (goles) con la jugadora X en cancha comparado con la jugadora X fuera de la cancha?*"
- En cancha: Primer partido X jugó 60 minutos, pierden 0-1. Segundo partido X entró a los 60' cuando ganaban 3-0 y el partido termina 5-0 =>  $((-1/60) + (2/30)) \times 90 = +4.5$  gpp
- Fuera de cancha: Primer partido X no estuvo 30 minutos, y en el segundo 60 minutos =>  $((0/30) + (3/60)) = +4.5$  gpp
- $(+4.5) - (+4.5) = 0$  ("el equipo juega igual conmigo que sinmigo")



# ¿Qué puedo hacer con los datos que tengo?

## Nivel 1: Score +/-

- **¿El problema?**
  - Pocos goles
  - Baja rotación de jugadoras
- **¿La solución?**
  - +/- con GPIV (Kharrat et al. 2017)
  - Regularized Adjusted Plus-Minus (Rosenbaum 2004), en que pasa a ser un problema de regresión lineal

# Modelamiento de futbolistas

## Nivel 2

# TAG -> LOG

YouTubeCoder by FC Python

Enter YouTube URL  Change Video (Deletes Events)  Change Pitch

Share

Watch on YouTube

Team Player Event Mins Secs X Y X2 Y2

Export CSV  Delete last

Home Away Shot Shot Assist Dribble Cross Pass

Edit Tags  Add Tag

(a)

Warnings Goals-Offs Teams Barcelona Atlético Ma...

Pitch Buttons Matrix Players Module Settings

00:11:03.95

Time	Player	Event	Mins	Secs	X	Y	X2	Y2	
01	24 J.Giménez								
01	18 JesúsGámez								
01	22 T.Parrey								
08	6 Koke								
01	17 Saul								
06	21 Y.FerreiraCa								
00	14 Galx								
0	3 Moyá								
03	3 FilipeLuis								
09	12 A.Fernández								
08	2 D.Godín								
0	11 J.Martínez								
01	7 A.Graizmann								

zoom - 100 %

(b)

Warnings Goals-Offs Teams Barcelona Atlético Ma...

Pitch Buttons Matrix Players Module Settings

Warnings Goals-Offs Teams Barcelona Atlético Ma...

Pitch Buttons Matrix Players Module Settings

Player	Team	Position	Color
1	FCB	Goalkeeper	Green
2	FCB	Defender	Red
3	FCB	Defender	Red
4	FCB	Defender	Red
5	FCB	Defender	Red
6	FCB	Midfielder	Blue
7	FCB	Midfielder	Blue
8	FCB	Midfielder	Blue
9	FCB	Forward	Blue
10	FCB	Forward	Blue
11	FCB	Forward	Blue
12	FCB	Forward	Blue
13	FCB	Forward	Blue
14	FCB	Forward	Blue
15	FCB	Forward	Blue
16	FCB	Forward	Blue
17	FCB	Forward	Blue
18	FCB	Forward	Blue
19	FCB	Forward	Blue
20	FCB	Forward	Blue
21	FCB	Forward	Blue
22	FCB	Forward	Blue
23	FCB	Forward	Blue
24	FCB	Forward	Blue
25	FCB	Forward	Blue
26	FCB	Forward	Blue
27	FCB	Forward	Blue
28	FCB	Forward	Blue
29	FCB	Forward	Blue
30	FCB	Forward	Blue
31	FCB	Forward	Blue
32	FCB	Forward	Blue
33	FCB	Forward	Blue
34	FCB	Forward	Blue
35	FCB	Forward	Blue
36	FCB	Forward	Blue
37	FCB	Forward	Blue
38	FCB	Forward	Blue
39	FCB	Forward	Blue
40	FCB	Forward	Blue
41	FCB	Forward	Blue
42	FCB	Forward	Blue
43	FCB	Forward	Blue
44	FCB	Forward	Blue
45	FCB	Forward	Blue
46	FCB	Forward	Blue
47	FCB	Forward	Blue
48	FCB	Forward	Blue
49	FCB	Forward	Blue
50	FCB	Forward	Blue
51	FCB	Forward	Blue
52	FCB	Forward	Blue
53	FCB	Forward	Blue
54	FCB	Forward	Blue
55	FCB	Forward	Blue
56	FCB	Forward	Blue
57	FCB	Forward	Blue
58	FCB	Forward	Blue
59	FCB	Forward	Blue
60	FCB	Forward	Blue
61	FCB	Forward	Blue
62	FCB	Forward	Blue
63	FCB	Forward	Blue
64	FCB	Forward	Blue
65	FCB	Forward	Blue
66	FCB	Forward	Blue
67	FCB	Forward	Blue
68	FCB	Forward	Blue
69	FCB	Forward	Blue
70	FCB	Forward	Blue
71	FCB	Forward	Blue
72	FCB	Forward	Blue
73	FCB	Forward	Blue
74	FCB	Forward	Blue
75	FCB	Forward	Blue
76	FCB	Forward	Blue
77	FCB	Forward	Blue
78	FCB	Forward	Blue
79	FCB	Forward	Blue
80	FCB	Forward	Blue
81	FCB	Forward	Blue
82	FCB	Forward	Blue
83	FCB	Forward	Blue
84	FCB	Forward	Blue
85	FCB	Forward	Blue
86	FCB	Forward	Blue
87	FCB	Forward	Blue
88	FCB	Forward	Blue
89	FCB	Forward	Blue
90	FCB	Forward	Blue
91	FCB	Forward	Blue
92	FCB	Forward	Blue
93	FCB	Forward	Blue
94	FCB	Forward	Blue
95	FCB	Forward	Blue
96	FCB	Forward	Blue
97	FCB	Forward	Blue
98	FCB	Forward	Blue
99	FCB	Forward	Blue
100	FCB	Forward	Blue

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: playeRank (wyscout LOG), 2019

- Pappalardo et al. tomaron los datos de 5 ligas más importantes de Europa:
  - Crearon un vector de **76** tipos de eventos y/o propiedades a partir de los datos que entregaba wyscout
  - Se lo pasaron a un computador: *"resuélveme el problema de optimización de, dadas esas 76 propiedades, cuál de ellas me suma o resta para ganar un partido" ... "para eso usa los datos de todos los partidos de las 5 ligas"*.
  - El computador entrega una *"evaluación"* con la *"importancia"* de cada una de ese tipo de evento en pos de ganar el partido, esa evaluación se llama **peso**

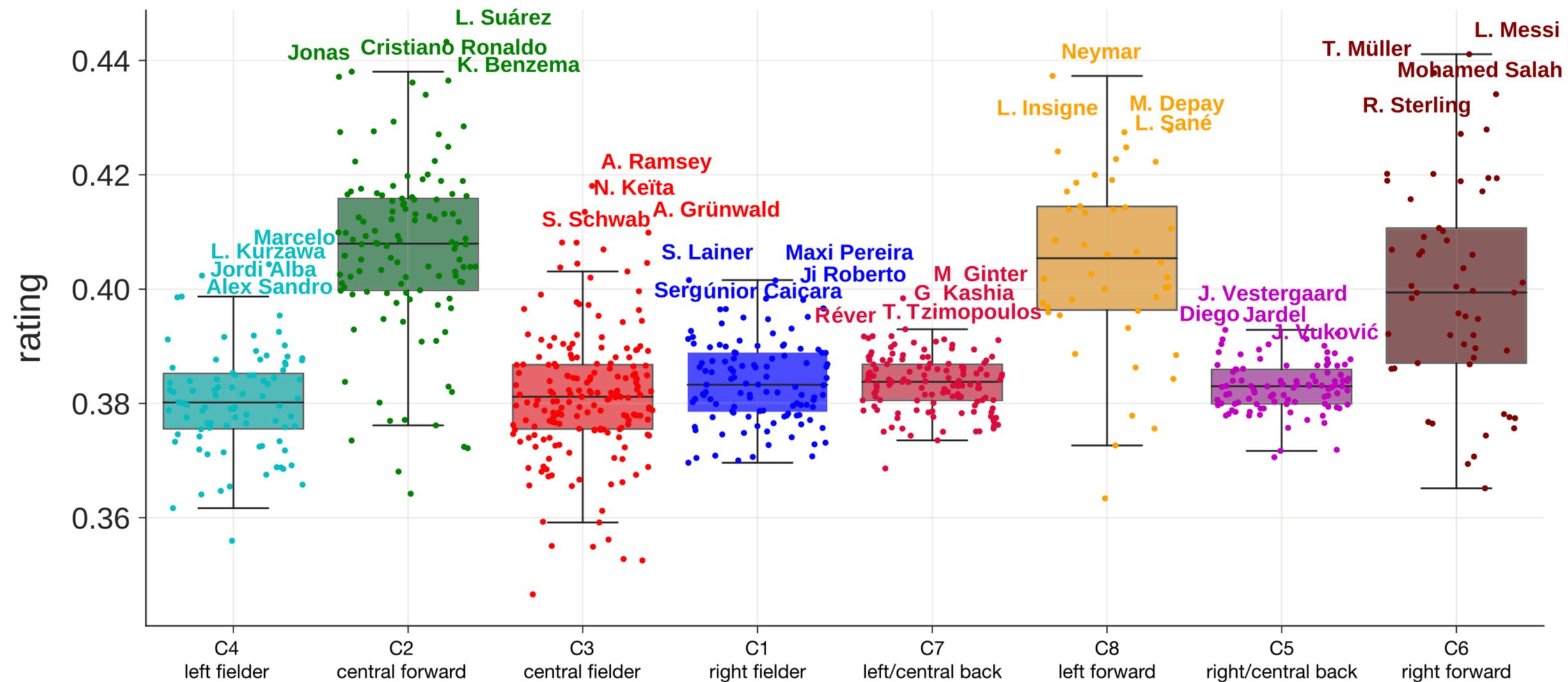
# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: playeRank (wyscout LOG), 2019

- Pappalardo et al. tomaron los datos de 5 ligas más importantes de Europa:
  - Para cada jugador a evaluar se toma todas sus acciones durante el partido, cada una de ellas se multiplica por su "peso" y se suma el total (por ejemplo, una tarjeta roja penaliza con -0.7, por decir un valor)
  - Se agrega (suma, promedia, pondera) la evaluación por partido en toda la temporada
  - Se obtiene el **playeRank** que además evalúa aproximadamente la toma de decisiones del jugador a lo largo del tiempo
  - **NOTA:** Jugadoras top no son las que sacan el mayor playeRank en un partido, son las que consistentemente sacan buen playeRank durante el torneo

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: playeRank (wyscout LOG), 2019



# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: Markov (Sarah Rudd, 2011)

- **El fútbol es una cadena de pases**
  - ... que termina con un cambio de posesión**
  - ... con un foul (cambio de posesión)**
  - ... con la pelota fuera (cambio de posesión)**
  - ... o con un gol**

# ¿Qué puedo hacer con los datos que tengo?

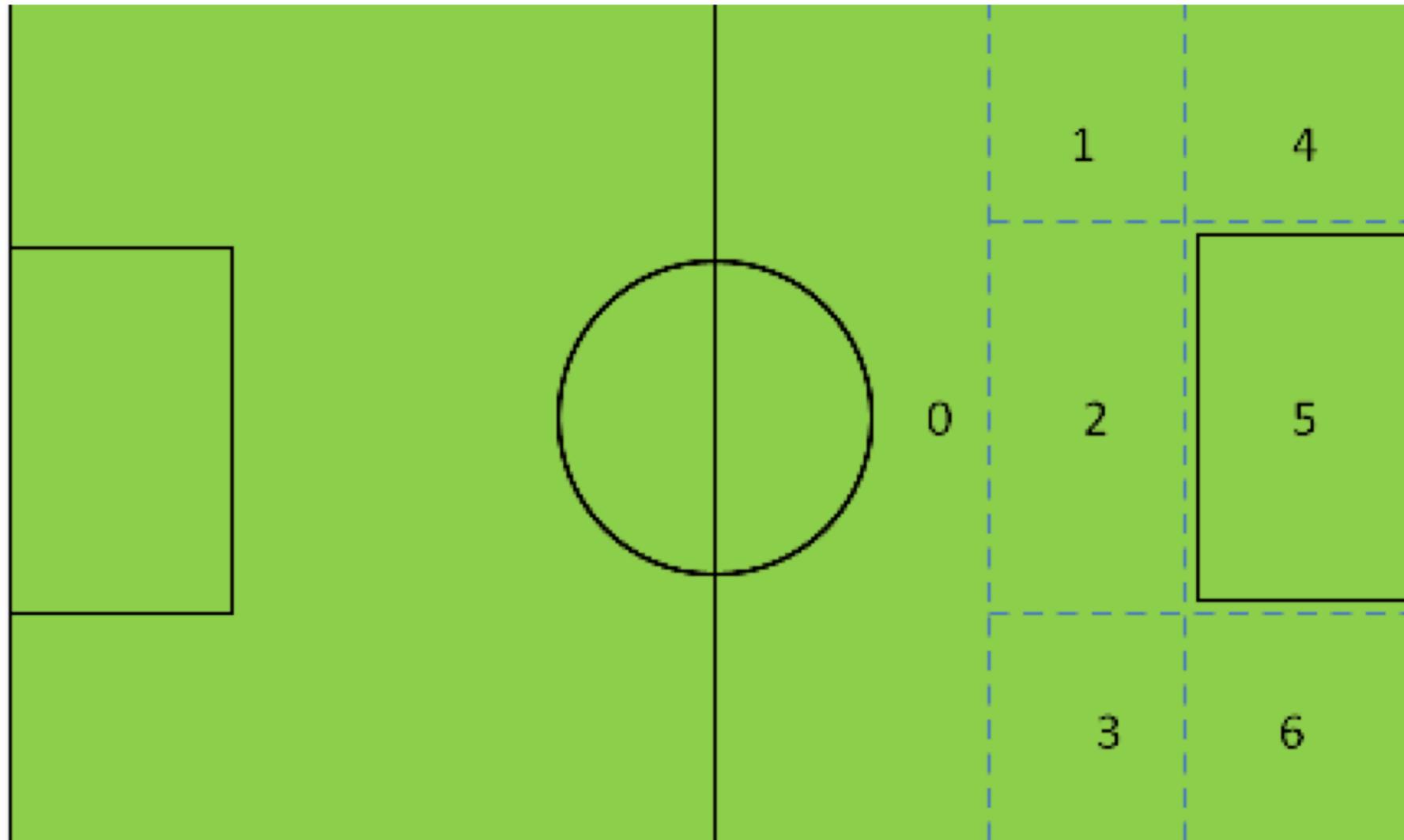
## Nivel 2: Markov (Sarah Rudd, 2011)

- Existen 7 tipos de jugada de "inicio" de juego:
  - 1) Penal
  - 2) Corner corto
  - 3) Corner largo
  - 4) Tiro libre corto
  - 5) Tiro libre largo
  - 6) Lateral corto
  - 7) Lateral largo

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: Markov (Sarah Rudd, 2011)

- Existen 7 zonas de juego:



# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: Markov (Sarah Rudd, 2011)

- **Existen entonces 39 estados:**
  - 1) **2 "absorbentes" (de donde no se sale): GOL y FIN POSESION**
  - 2) **7 jugadas de inicio**
  - 3) **30 estados definidos por localización y acción (centros, tiros)**

**= 39 estados**

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: Markov (Sarah Rudd, 2011)

- Esta matriz nos da la probabilidad de pasar de un estado a otro

	End of Pos.	Goal	$S_1$	...	$S_n$
End of Pos.	1	0	0	...	0
Goal	0	1	0	...	0
$S_1$	.5	.02	.05	...	.03
...	⋮	⋮	⋮	⋱	⋮
$S_n$	.6	.15	.02	...	.4

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: Markov (Sarah Rudd, 2011)

- Existen entonces 39 estados:
  - 1) 2 "absorbentes" (de donde no se sale): GOL y FIN POSESION
  - 2) 7 jugadas de inicio
  - 3) 30 estados definidos por localización y acción (centros, tiros)

= 39 estados
- Obtiene la  $P(\text{gol})$  a partir de cada estado, por ejemplo  
 $P(\text{GOL desde un PENAL}) = 0.71$

# ¿Qué puedo hacer con los datos que tengo?

## Nivel 2: Markov (Sarah Rudd, 2011)

- ¿Cómo se hace un score/ranking con eso?
- Se evalúa cada acción (pasa de estado x a estado y) de jugadora y se evalúa con la diferencia de la probabilidad de gol.

