A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains

Sarah Rudd

On Football Research and Consulting

What are we trying to solve?

- How valuable are certain game situations?

- How do these values vary across teams and what can we learn from that?

- Can we quantify how much a player contributes to creating good goal scoring opportunities?

Why is this hard?

- Hard to capture all of the information about game state; off the ball movement is just as important as on the ball movement

- Sparse data – we haven't seen all possible combinations of game situations, and those we have seen only have a few data points

- How do you divide up credit?

Markov Chains

- **What do they do?**
  - Model the likely outcomes after a number of iterations based on the probabilities of transitioning from one state to another

- **Why are they useful?**
  - They allow us to look at all the possible ways a possession can unfold
  - Absorption states means possessions of arbitrary lengths are handled nicely

- **Downside?**
  - Assume that the current state is independent to the previous state (i.e. it doesn't matter how we got here, the probabilities of moving to the next state are the same regardless of the past)

Dataset

- Touch-by-touch data provided by StatDNA
  - *(x,y)* coordinates
  - event type
  - defensive pressure
  - defensive "state"

- English Premier League
  - 2010/11 Season
  - 123 matches
    - Minimum of 11 matches per team
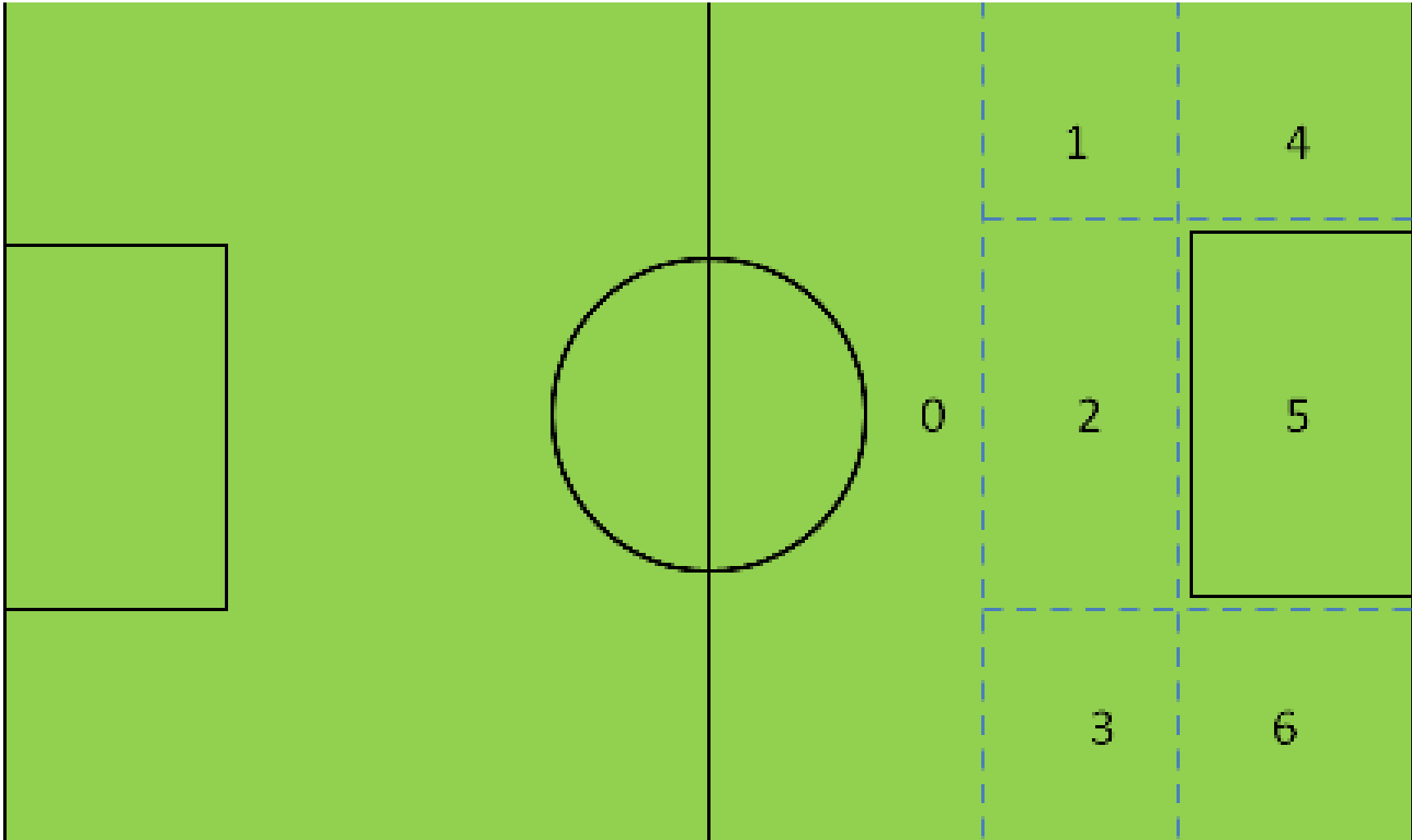   - ~100,000 "deliberate" actions or about 800 actions per match

States

- 2 absorbing states: Goal or End of Possession
- 7 set pieces
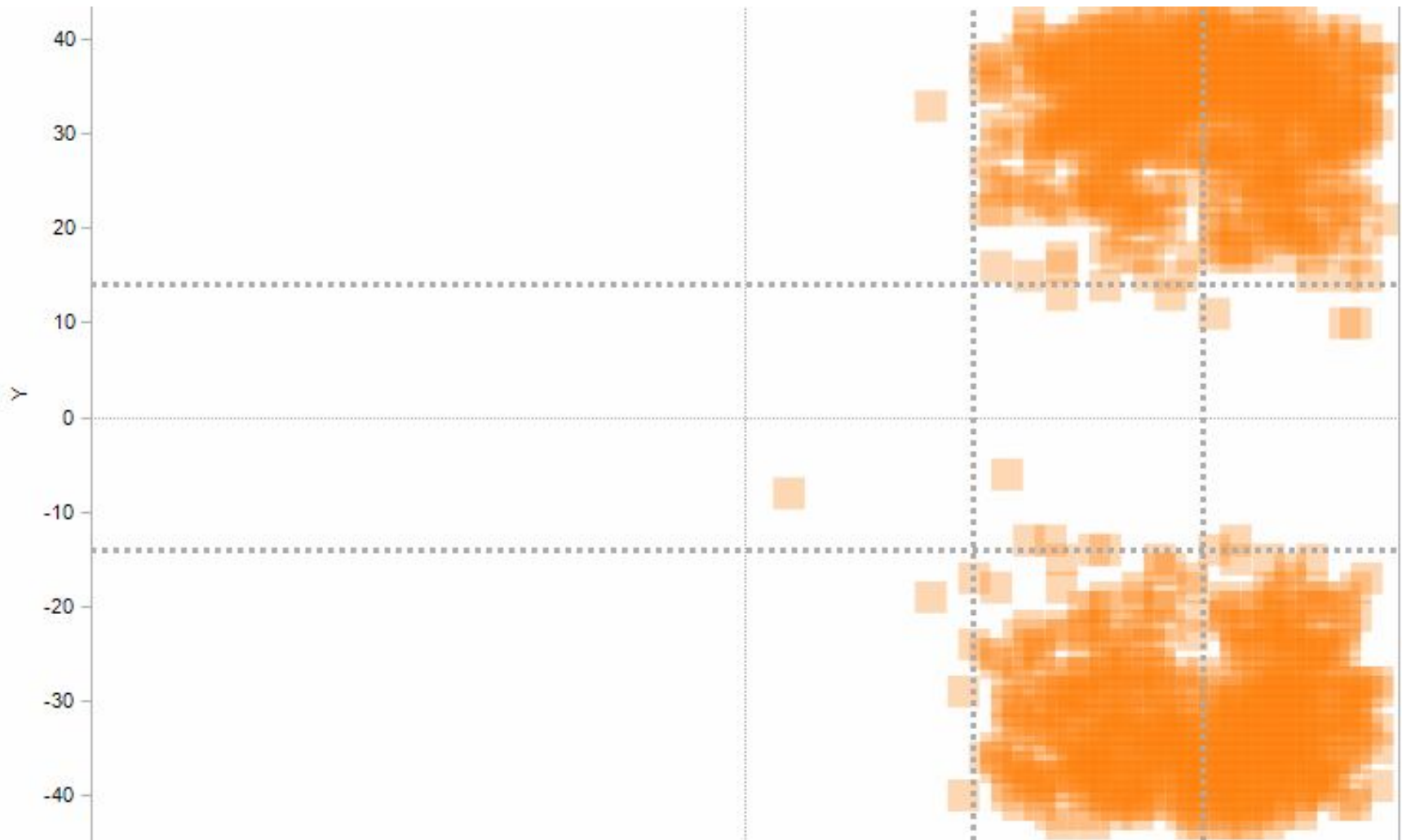- 30 states defined by zonal location and defensive state

39 total states

Set Pieces

- Penalty
- Short Corner
- Long Corner
- Short Free Kick
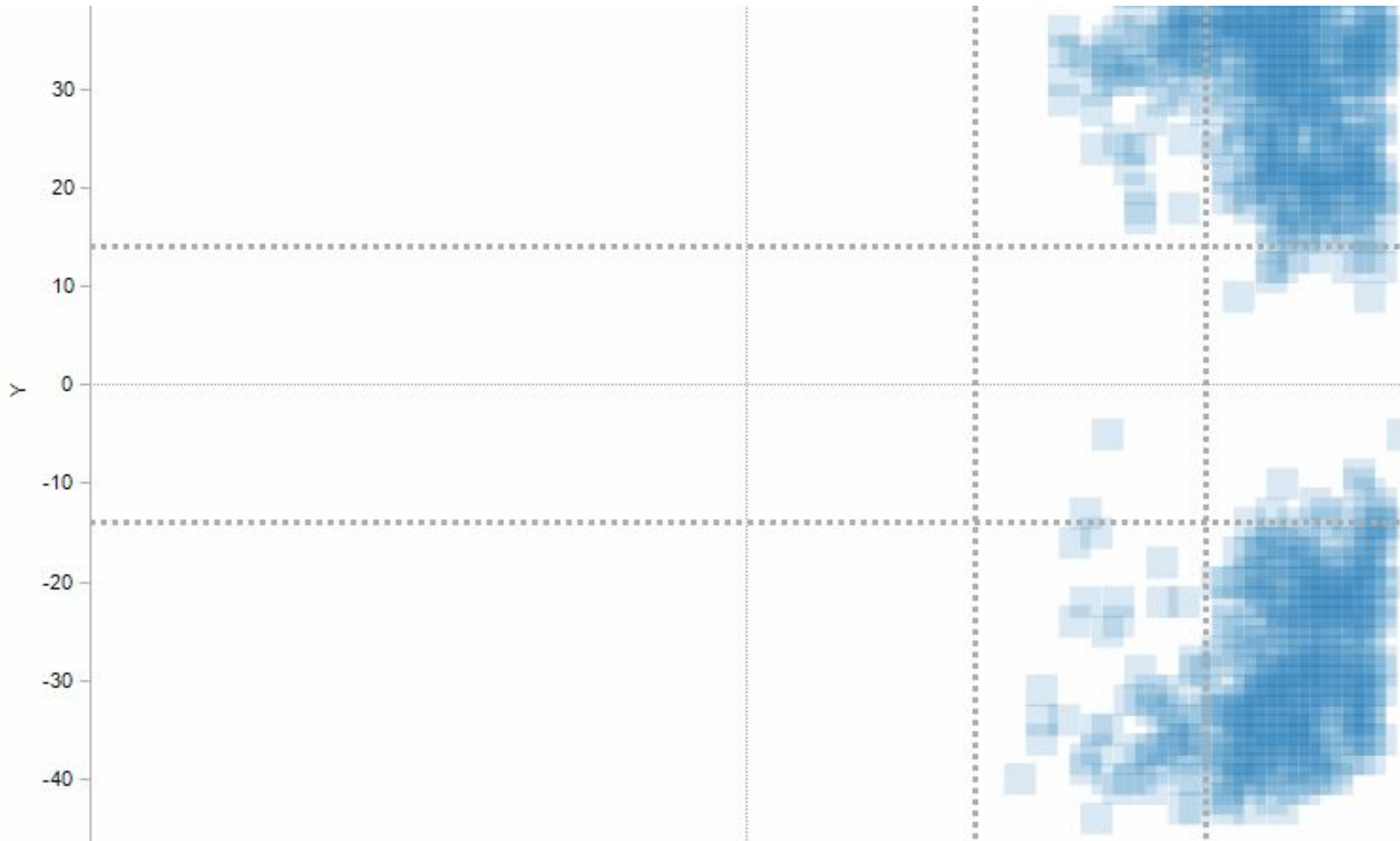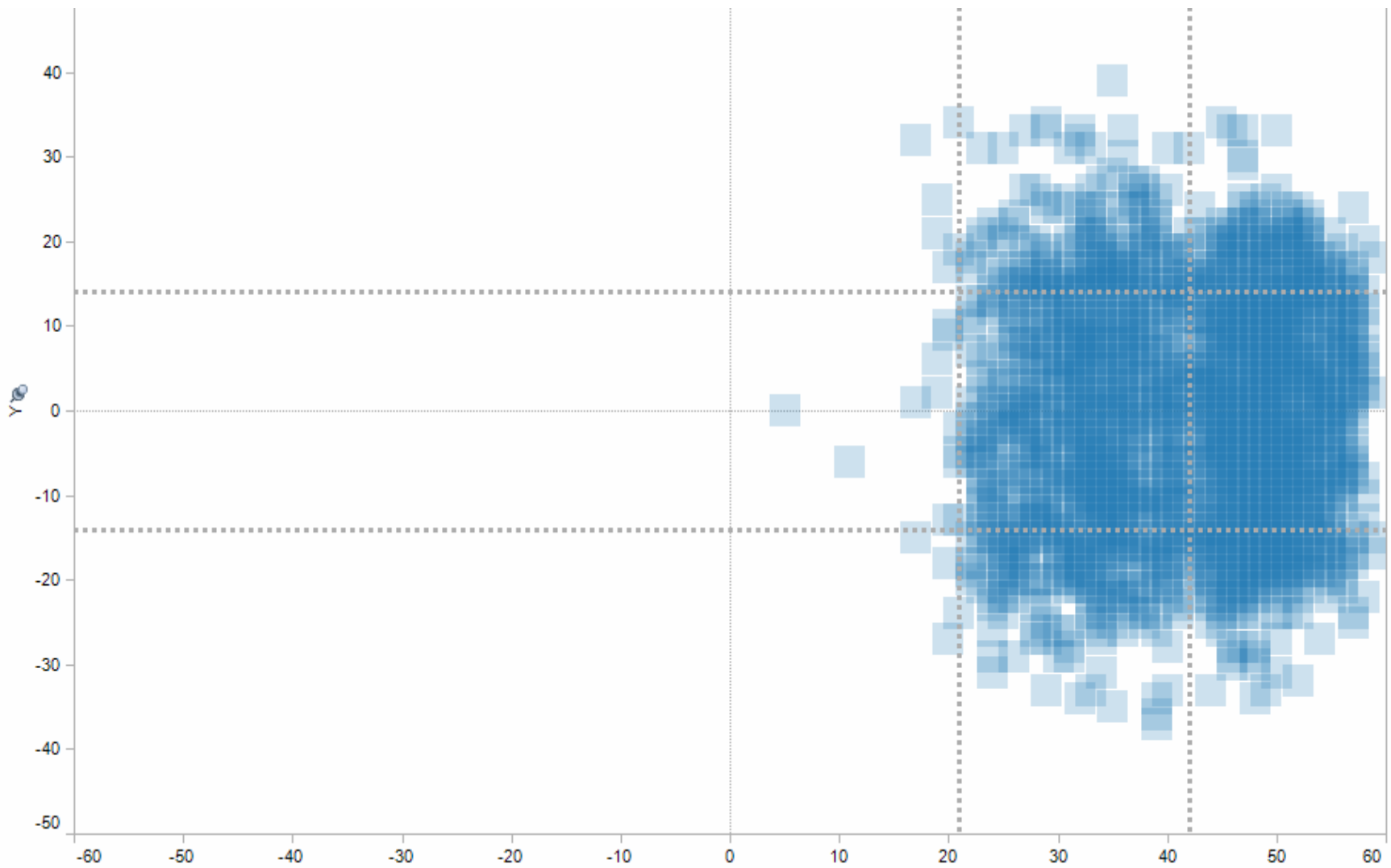- Long Free Kick
- Shallow Throw-in
- Deep Throw-in

Zones

Crosses (Pressure A)

Crosses (Pressure B)

Shots

Goals
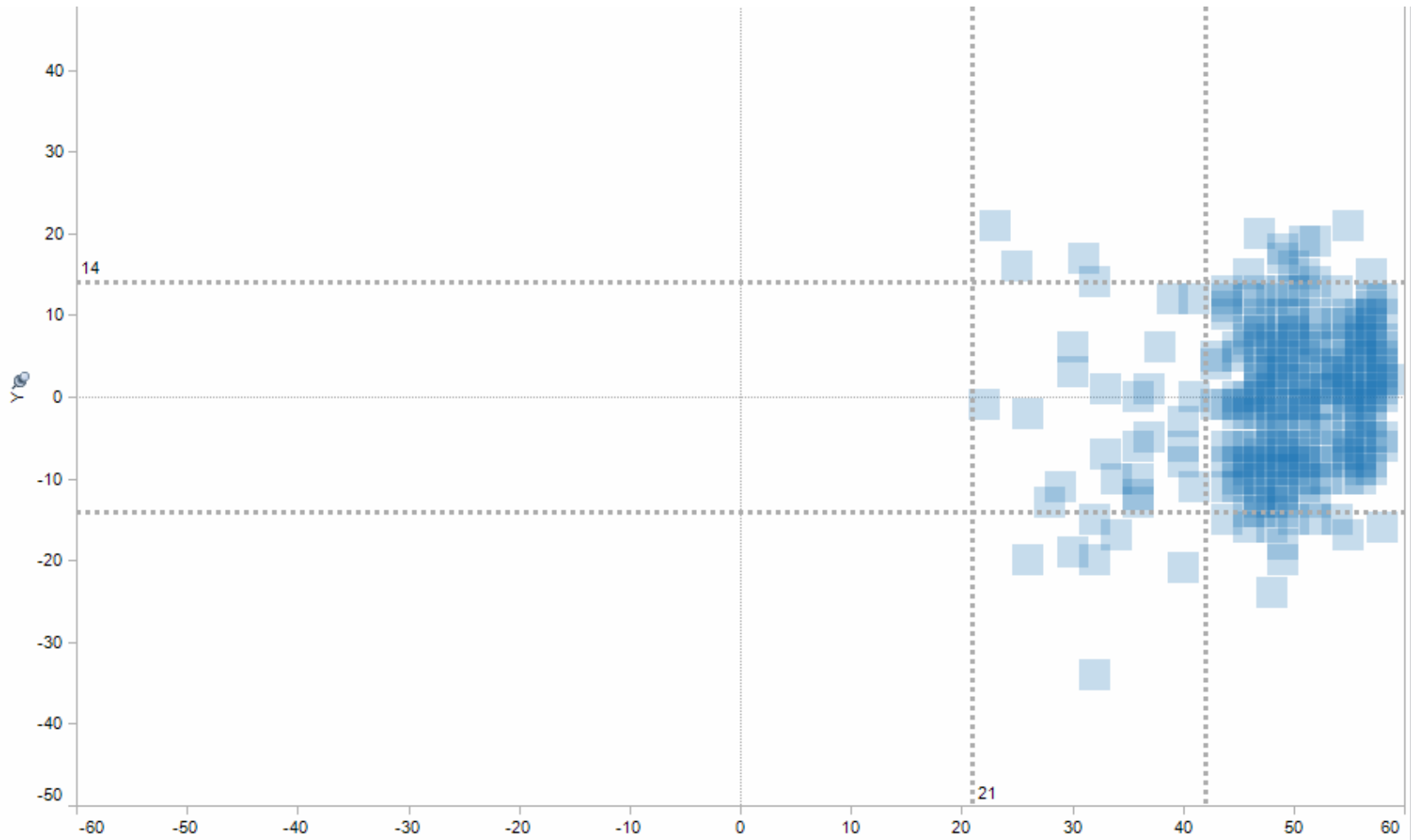
Definitions

- **Deliberate action** – any action where a player moves the ball in a controlled manner with an attempted outcome
  - Deliberate: pass, shot, dribble, etc.
  - Not deliberate: clearance, tackle, etc.

- **Possession** – a series of consecutive deliberate actions performed by one team, only interrupted by a deliberate action performed by the other team or the end of a half.

Transition Matrices

- Calculate the probability of moving from state S_a to S_b for all combinations of the 39 states

- Absorbing states are different, probability of remaining in same state is 1 and moving to another state is 0.
  - *Once you are there you are stuck!*

- Multiplying a transition matrix by itself will give you the probability of ending up in a given state after 1 iteration

- Multiplication can be repeated until probability of ending in an absorbing state converges (*n=20*)

Transition Matrix

$$
\begin{array}{c} \\ \text{End of Pos.} \\ \\ \text{Goal} \\ \\ S_1 \\ \\ \vdots \\ \\ S_n \end{array}
\begin{array}{cc}
\text{End of Pos.} & \text{Goal} & & S_1 & & & & S_n \\
\end{array}
$$

	End of Pos.	Goal	S_1		S_n
End of Pos.	1	0	0	$\cdots$	0
Goal	0	1	0	$\cdots$	0
S_1	.5	.02	.05	$\cdots$	.03
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
S_n	.6	.15	.02	$\cdots$	.4

Validation – expected vs. actual

Monte Carlo Bootstrapping

- 1000 samples with replacement

$y = 0.9417x + 0.0134$
$R^2 = 0.99034$
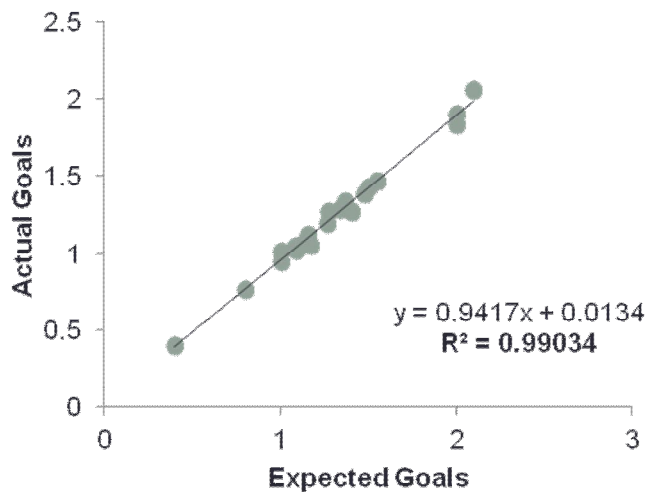
K-Fold Validation

- 5 folds, 4 training and 1 one evaluation

$y = 0.9126x - 0.0832$
$R^2 = 0.97564$
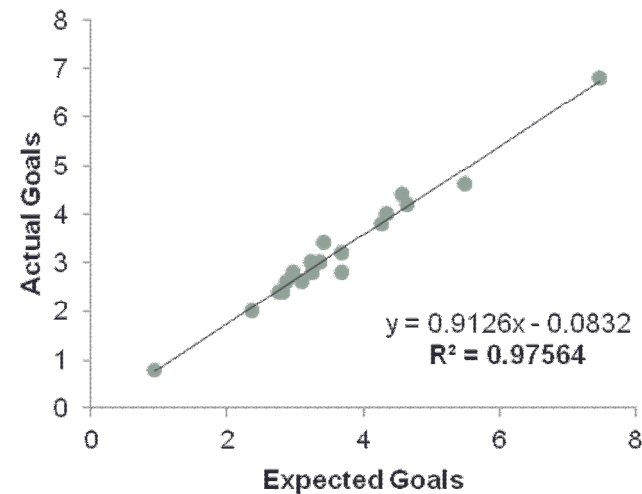
Comparing P(Goal)

- Column are teams ordered by final standing

- Rows are P(Goal) for each state

- As you move lower down the table, teams find it harder to score

- Notable exceptions are Manchester City (3rd, underperform offensively) and Wolves (17th, overperform offensively)
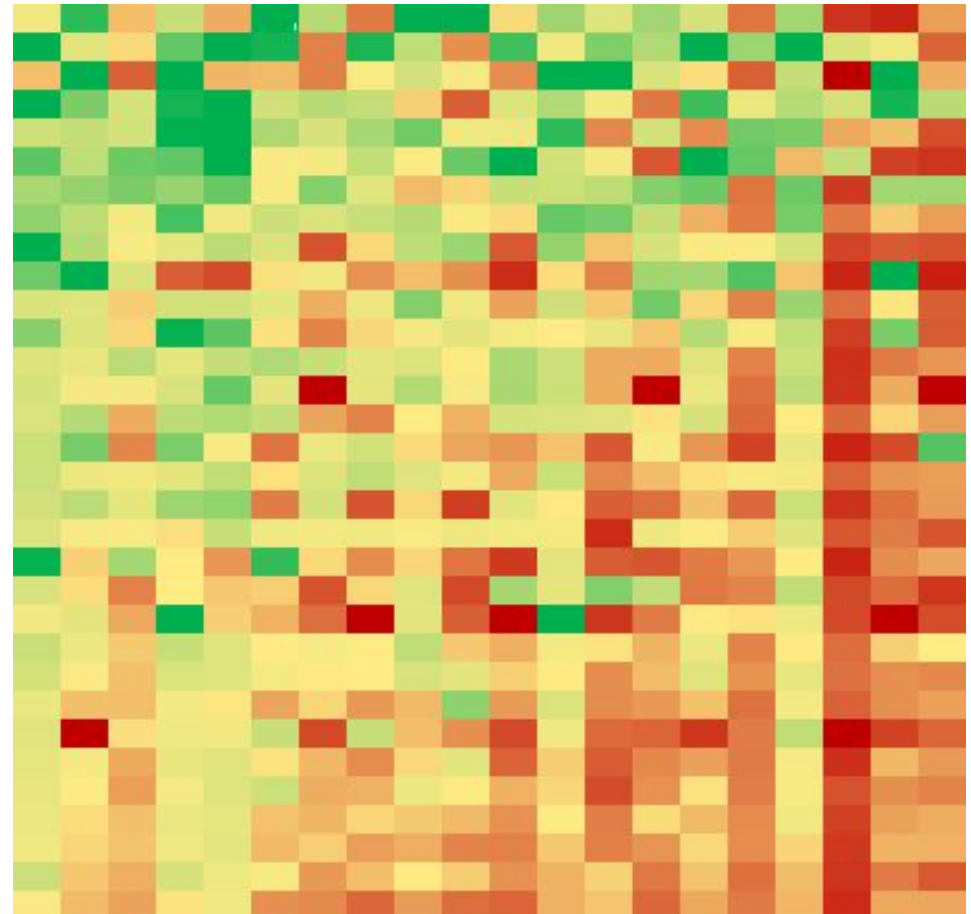
Results – Set Pieces

State	P(Goal)
Penalty	71.55%
Long Corner	2.39%
Short Corner	1.67%
Long Free Kick	1.67%
Shallow Throw in	1.46%
Deep Throw in	1.09%
Short Free Kick	1.08%

Corners

Offensive Probabilities
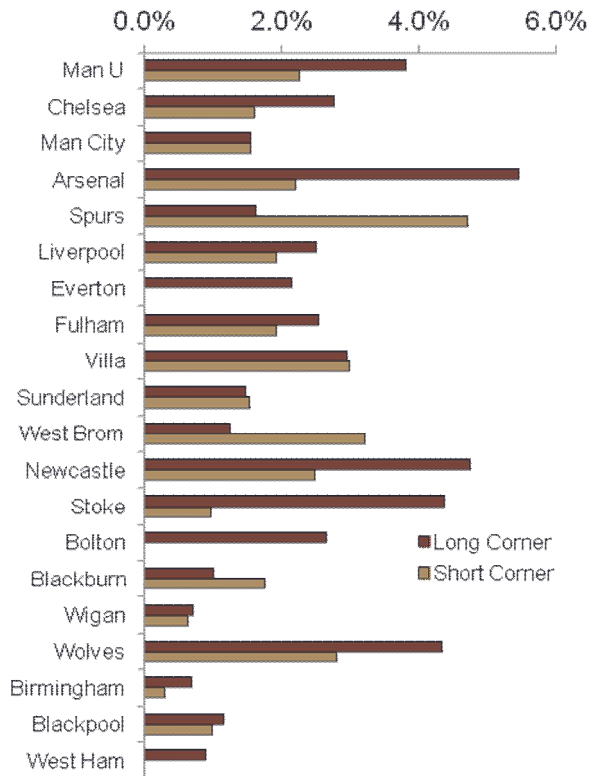
Defensive Probabilities
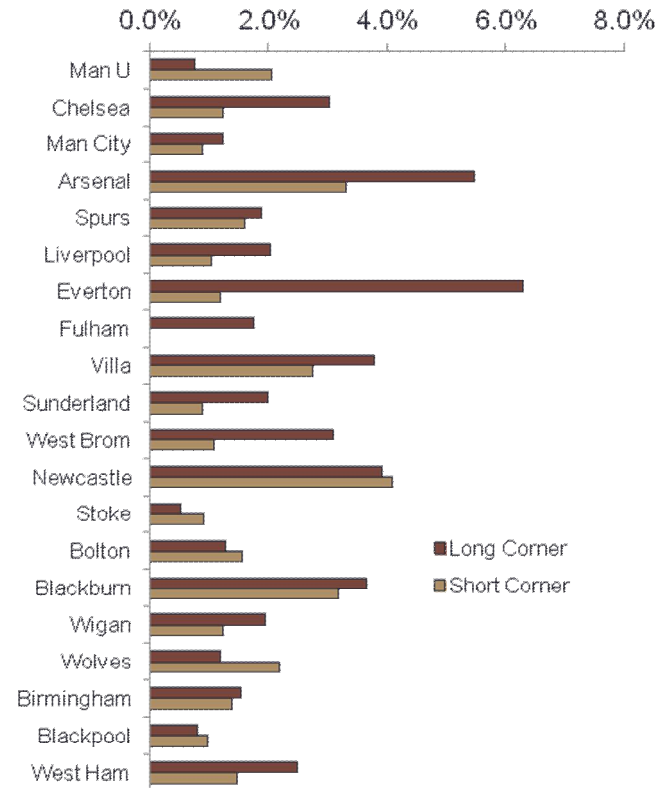
Counter Attacks
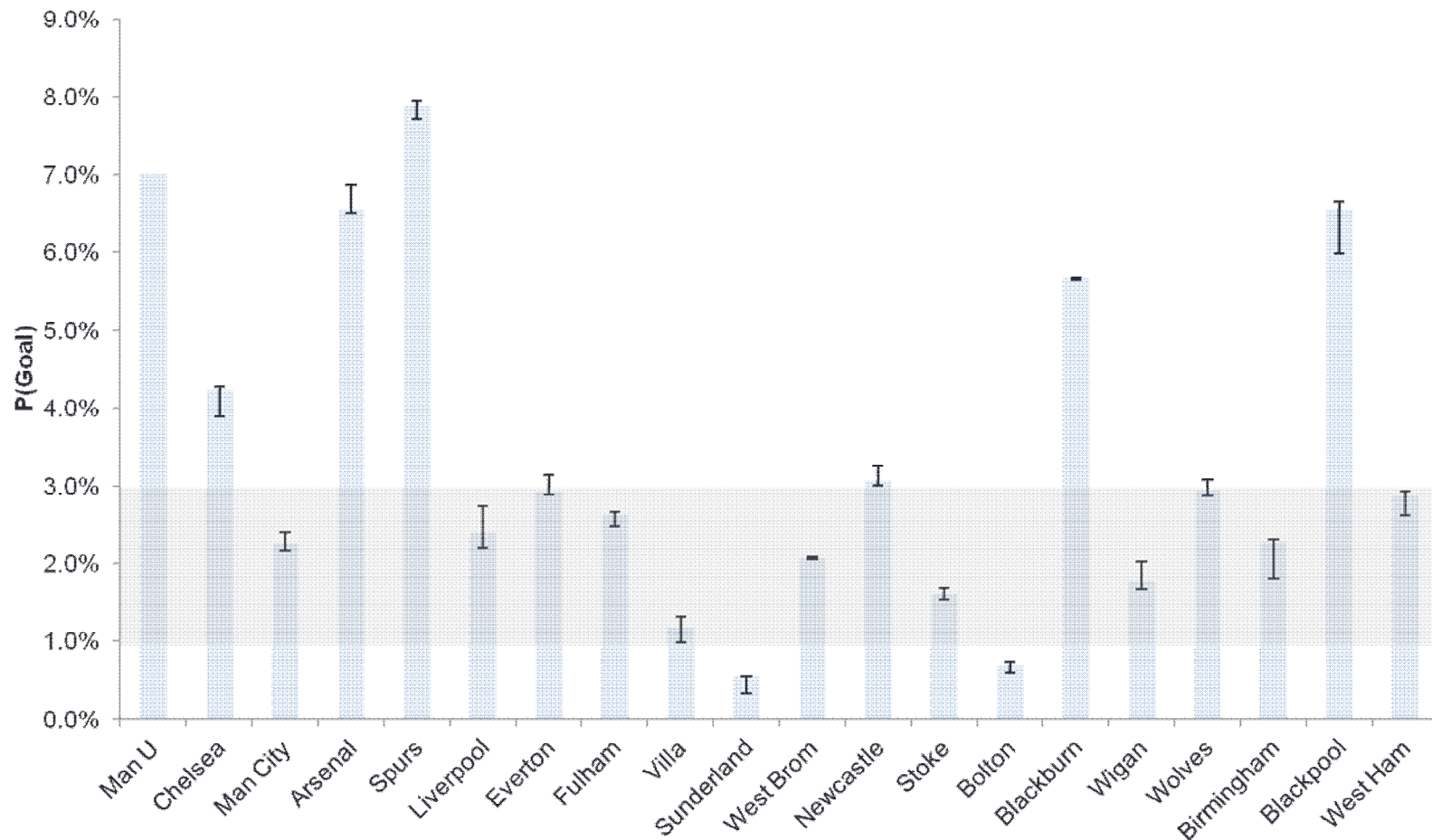
Now for the good stuff…

Not all passes are created equal

- Existing metrics don't take into account context of game state
  - Completed passes: back-pass, square-pass, through-ball that puts teammate 1-on-1 with keeper are all weighted equally
  - Goals: all are weighted equally, no matter how easy they were to score
  - Missed opportunities could still show up positively in metrics (i.e. a saved penalty could be considered a shot on target)

Weight each action with incremental improvement of P(Goal)

How does it work?

Player 1
State A
P(Goal)=
0.25

Player 2
State B
P(Goal)=
0.17

Player 3
State C
P(Goal)=
0.28

State:
Goal
Scored
P(Goal)=
1

Player 1:
-0.08

Player 2:
+0.11
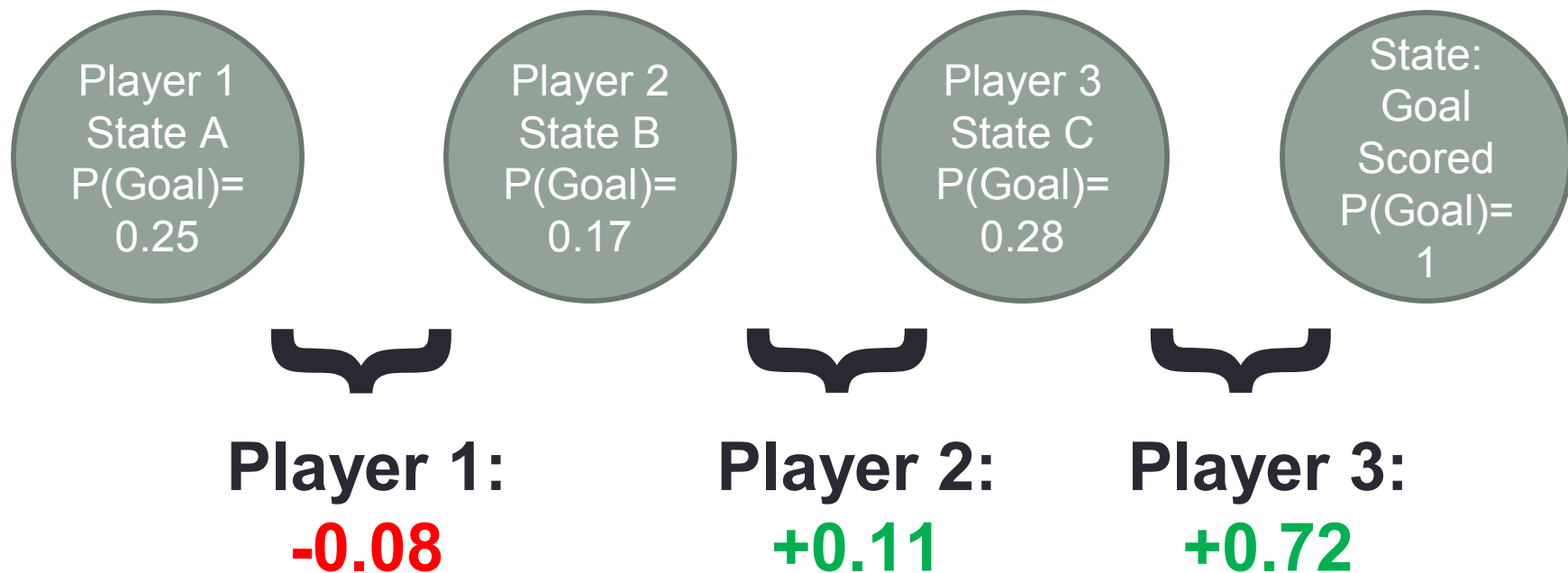
Player 3:
+0.72

Another example

Player 1
State A
P(Goal)=
0.15

Player 1
Earns
Penalty
P(Goal)=
0.71

Player 2
Takes
Penalty
P(Goal)=
0.71

State:
Penalty
Missed
P(Goal)=
0

Player 1:
+0.56

Player 2:
-0.71

Player 1 is rewarded for earning the penalty and Player 2 is heavily penalized for missing it.

Top Performers

• Top performers are:
- Tim Cahill
- Yaya Toure
- Cesc Fabregas

• Jordan Henderson in top 25
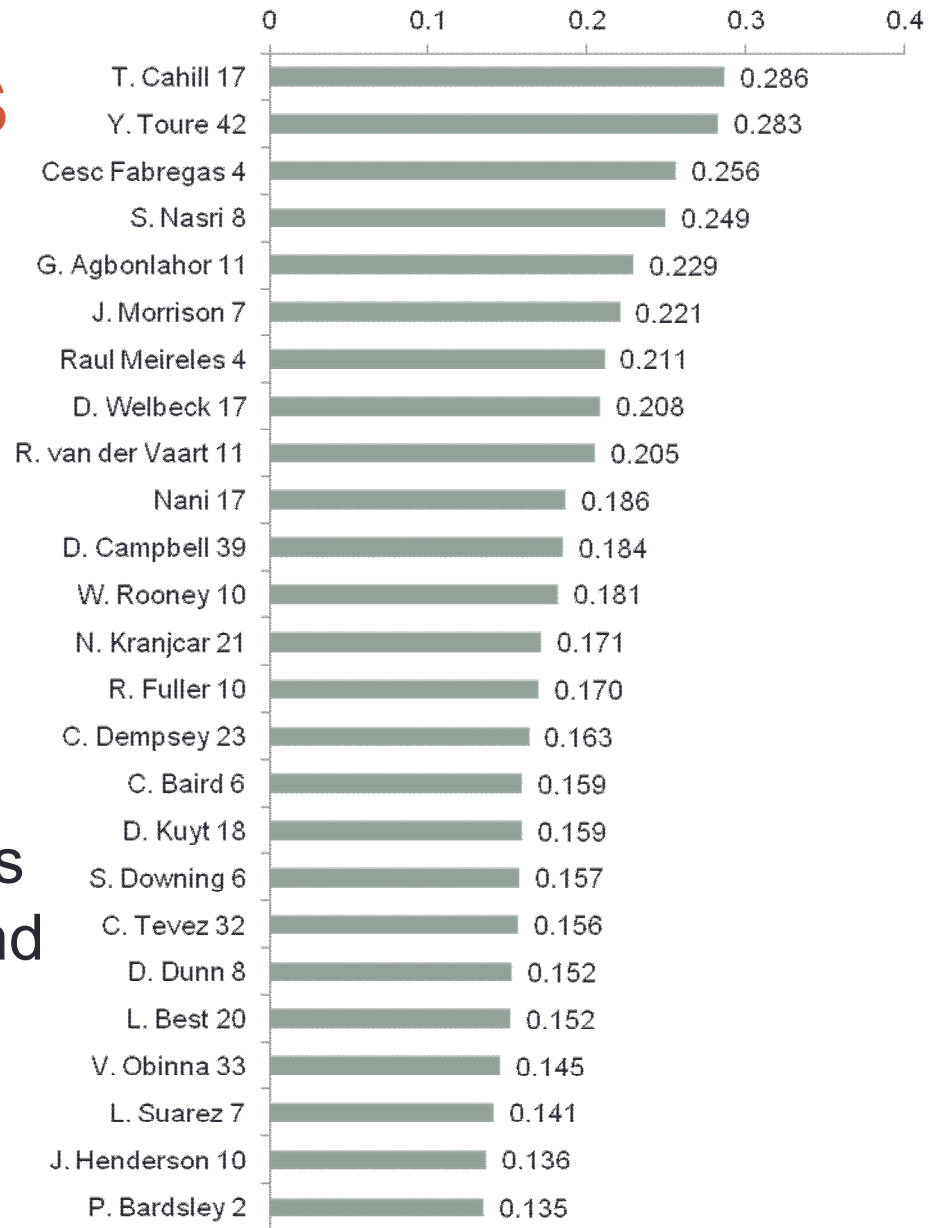
• Some surprises like James Morrison, Ricardo Fuller and Chris Baird

Player	Value
T. Cahill 17	0.286
Y. Toure 42	0.283
Cesc Fabregas 4	0.256
S. Nasri 8	0.249
G. Agbonlahor 11	0.229
J. Morrison 7	0.221
Raul Meireles 4	0.211
D. Welbeck 17	0.208
R. van der Vaart 11	0.205
Nani 17	0.186
D. Campbell 39	0.184
W. Rooney 10	0.181
N. Kranjcar 21	0.171
R. Fuller 10	0.170
C. Dempsey 23	0.163
C. Baird 6	0.159
D. Kuyt 18	0.159
S. Downing 6	0.157
C. Tevez 32	0.156
D. Dunn 8	0.152
L. Best 20	0.152
V. Obinna 33	0.145
L. Suarez 7	0.141
J. Henderson 10	0.136
P. Bardsley 2	0.135

Worst Performers

- Lots of goal keepers, also strikers and defenders
- Poor Darren Bent
  - 1 goal was in sample set out of 17 he scored in the entire season
  - Had 19 opportunities where he had the ball with >10% chance of scoring (22% average) but only converted one
- Clichy/Young/Kolarov – poor crossers of the ball?
  - Can dig deeper into the data to identify which situations

Offensive Contribution	Player
-0.282	D. Bent 11
-0.195	P. Odemwingie 24
-0.176	G. Clichy 22
-0.167	K. Davies 14
-0.165	L. Fabianski 21
-0.154	D. N'Gog 24
-0.153	T. Howard 24
-0.145	M. Chamakh 29
-0.141	Gomes 1
-0.138	W. Hennessey 13
-0.136	P. Cech 1
-0.131	A. Kolarov 13
-0.127	A. Young 7
-0.126	L. Varney 16
-0.125	C. Berra 16
-0.125	R. Shawcross 17
-0.123	R. Dunne 5
-0.116	B. Hangeland 5
-0.114	E. van der Sar 1
-0.113	R. Green 1
-0.110	N. Ranger 30
-0.110	D. Duff 16
-0.109	G. Cahill 5
-0.108	P. Crouch 15
-0.108	C. Jerome 10

Offensive Contribution

Thank you for listening and special thanks to StatDNA for providing me with awesome data and this wonderful opportunity