Player Vectors:
Characterizing Soccer Players' Playing Style from Match Event Streams

Tom Decroos and Jesse Davis

September 19th, 2019

ECMLPKDD

Why can’t Dybala and Messi play in the same team?

Paulo Dybala

“It's difficult to play with Leo because we both play in the same position. I try to leave him his space but it's not easy. I have to adapt"

“Paulo plays in the same role as me with Juventus, and in general we play in the same position. We tried to play together but we realised that we tend to occupy the same area.”

Lionel Messi

What is playing style?

Our definition:
"A player's preferred area(s) on the field and
which actions he performs in each of these locations."

Our task:
Characterize playing style in a player vector that is

(1) Human-interpretable

(2) Suitable for data analysis

Outline

- Data and Challenges
- Building Player Vectors
- Experiments

Outline

- Data and Challenges
- Building Player Vectors
- Experiments

Event stream data describe events in a soccer match

Rich data source, BUT …

- Vendor-specific terminology

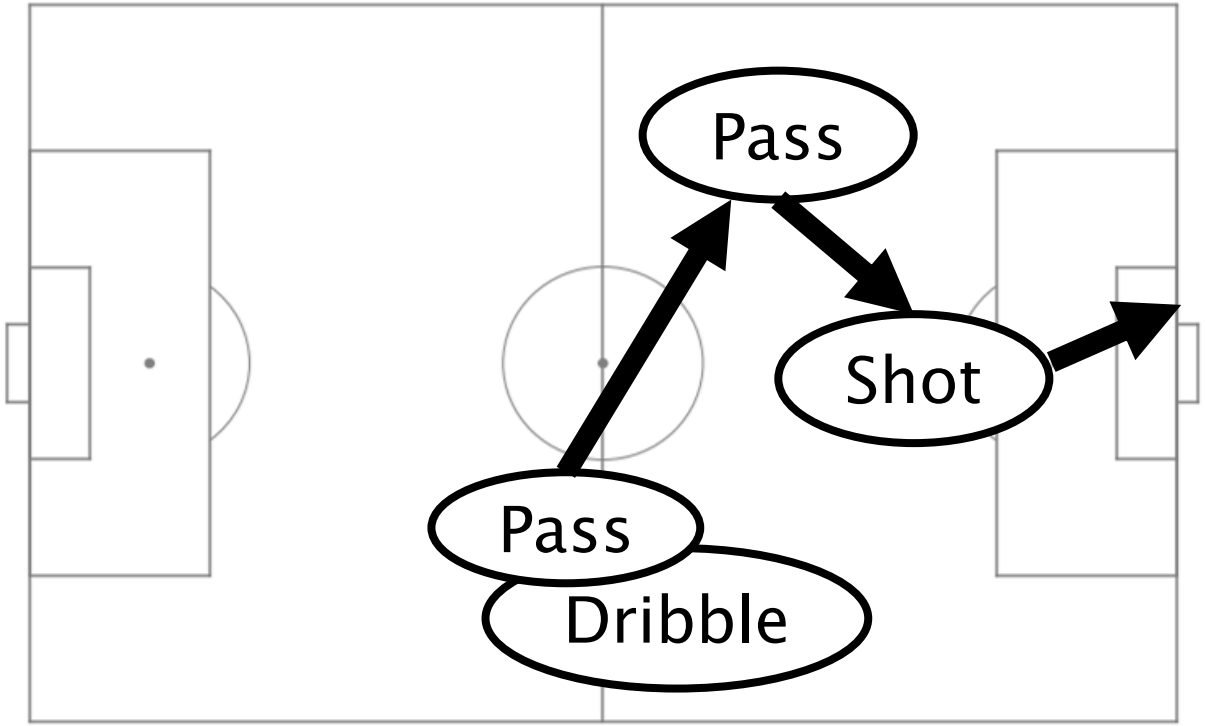
wyscout
opta STATSBOMB

- Useless events

Weather change

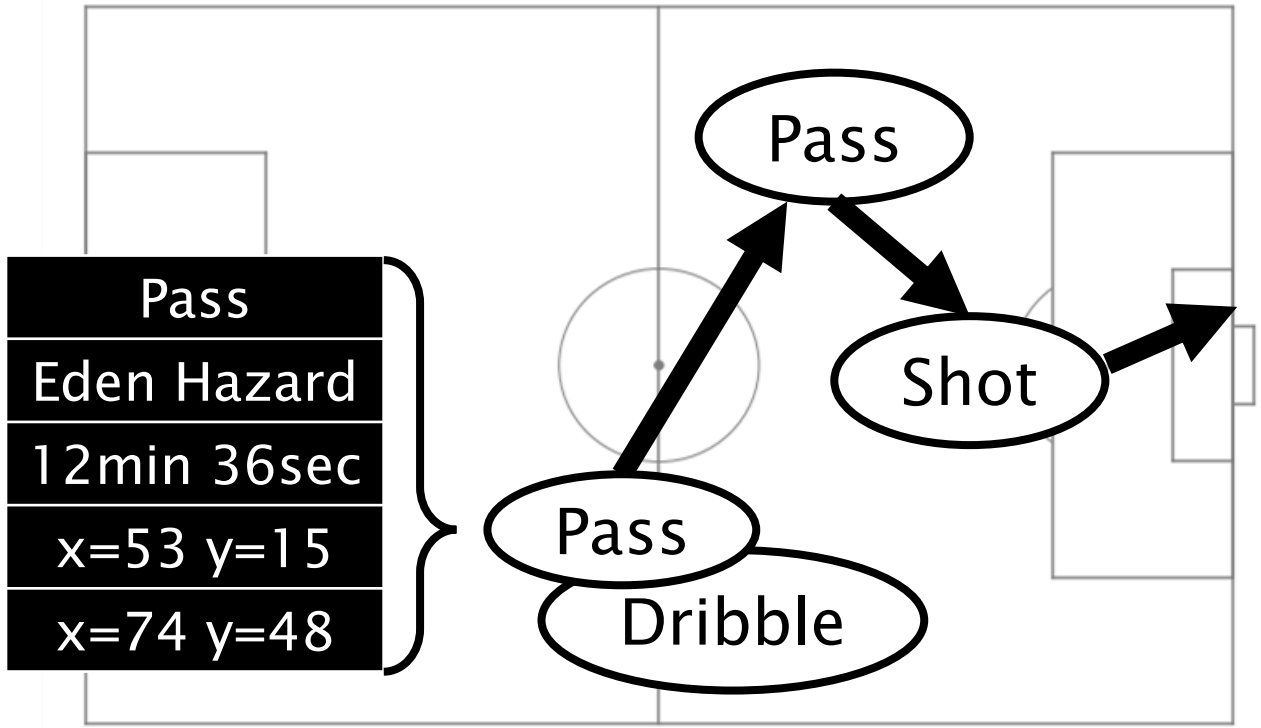
- Optional information snippets

```
{“deeply”:
  {“nested”:
    {“json”:
      {“dictionaries”}}}}
```



SPADL is a soccer match event stream data format designed to facilitate data analysis

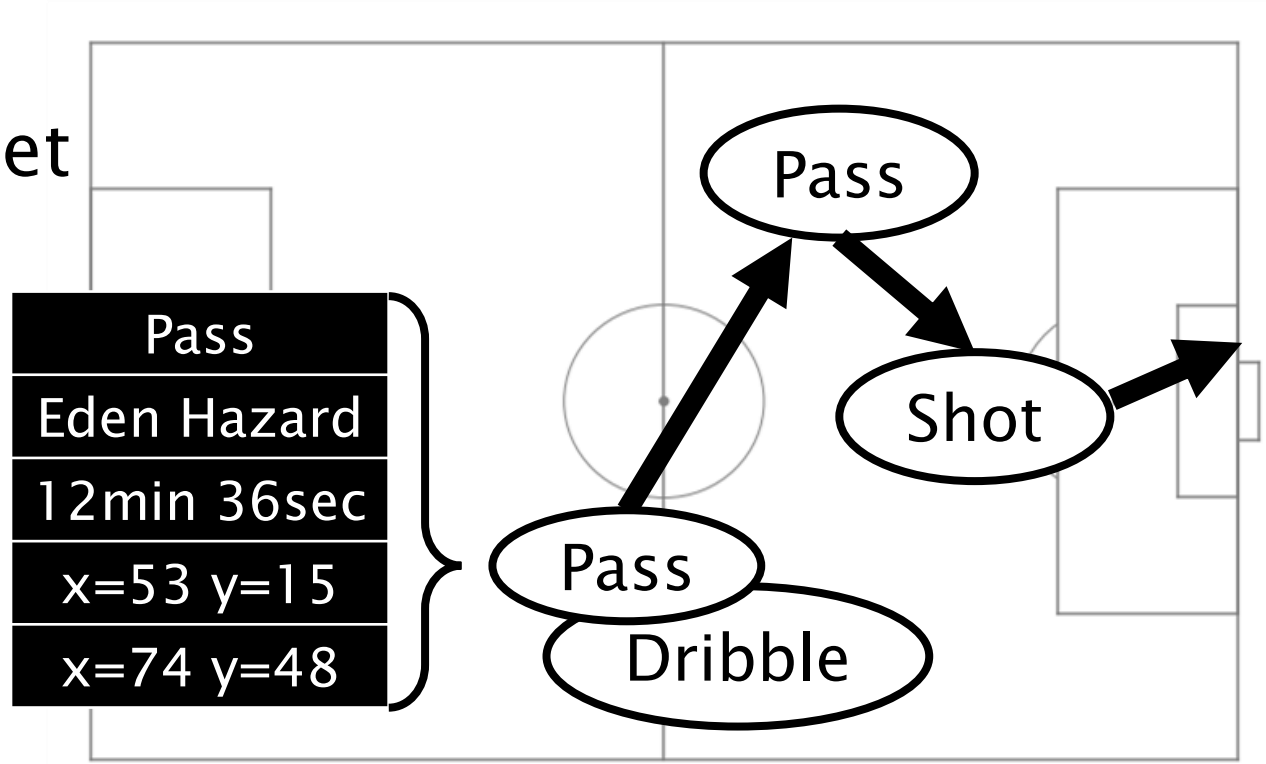
- Unifies data from various vendors
- On-the-ball actions
- Fixed attributes



Decroos et al. *Actions Speak Louder Than Goals: Valuing Player Actions in Soccer.* KDD 2019
https://github.com/ML-KULeuven/socceraction

Applying data science methods can be challenging

1. Categorical + continuous attributes
2. Variable length action set
3. No spatial repetition



Outline

- Data and Challenges
- Building Player Vectors
- Experiments

Building player vectors

Given: Event stream data of a player p

Do: Characterize player p's playing style in a vector that is human-interpretable and suitable for data analysis

Building player vectors in 4 steps

Given: Event stream data of a player p

Do: Characterize player p's playing style in a vector that is human-interpretable and suitable for data analysis

Approach:

1)

2)

3)

4)

Building player vectors in 4 steps

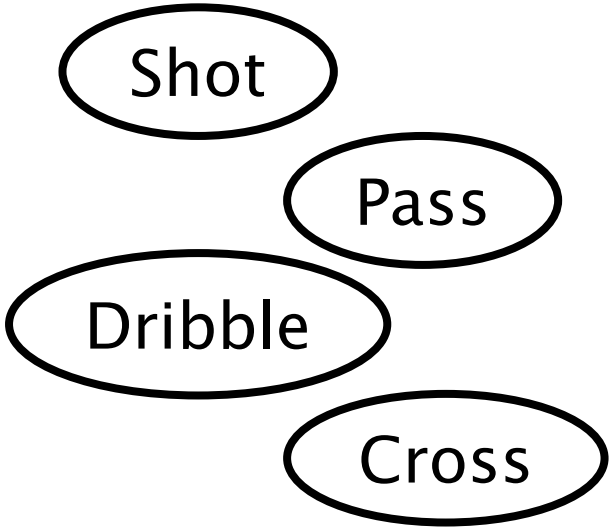
Given: Event stream data of a player p

Do: Characterize player p's playing style in a vector that is human-interpretable and suitable for data analysis

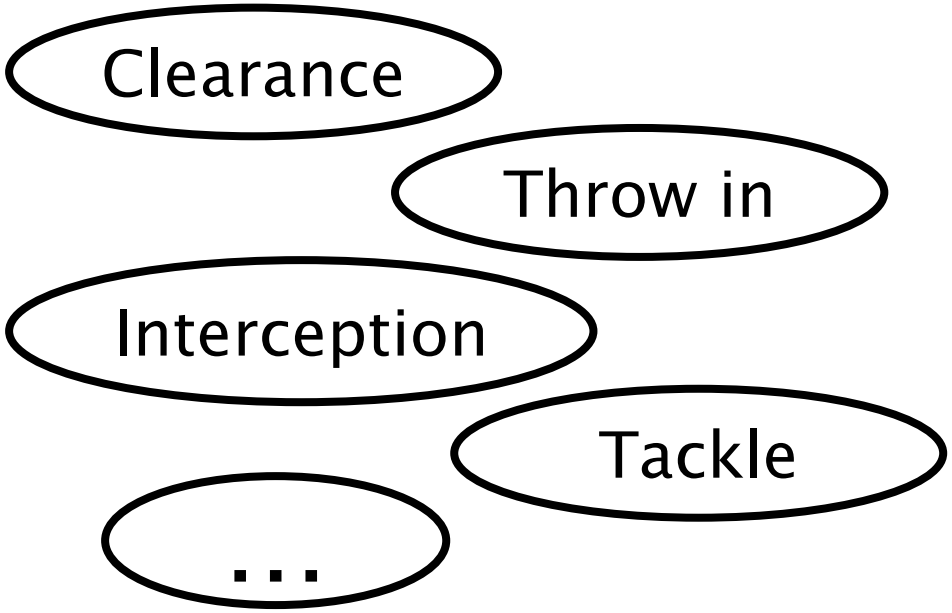
Approach:

1) Select relevant action types
2)
3)
4)

Relevant actions are offensive and occur during open play



Relevant



NOT relevant

Building player vectors in 4 steps

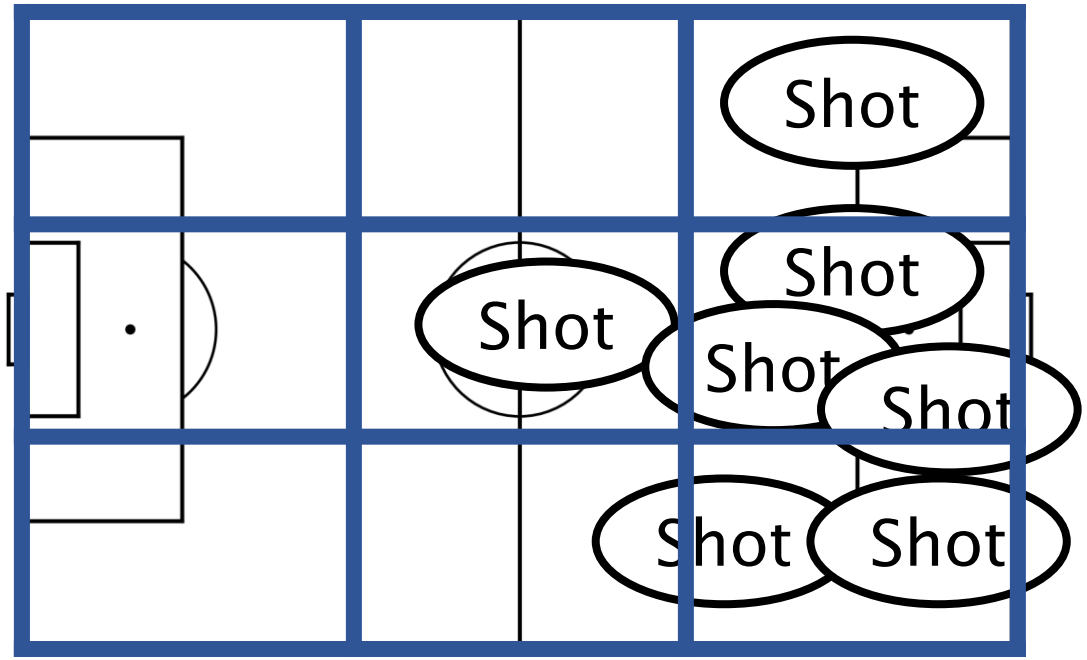
Given: Event stream data of a player p

Do: Characterize player p's playing style in a vector that is human-interpretable and suitable for data analysis

Approach:

1) Select relevant action types
2) Construct heatmaps per player per action type
3)
4)

Toy example: overlaying a grid and counting

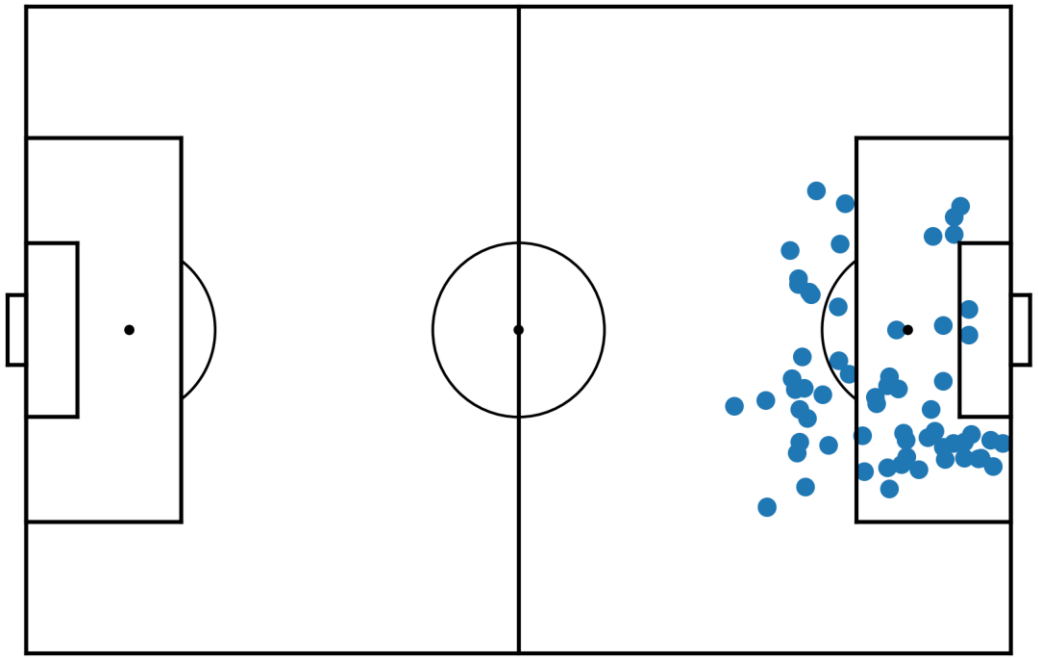


Scatterplot

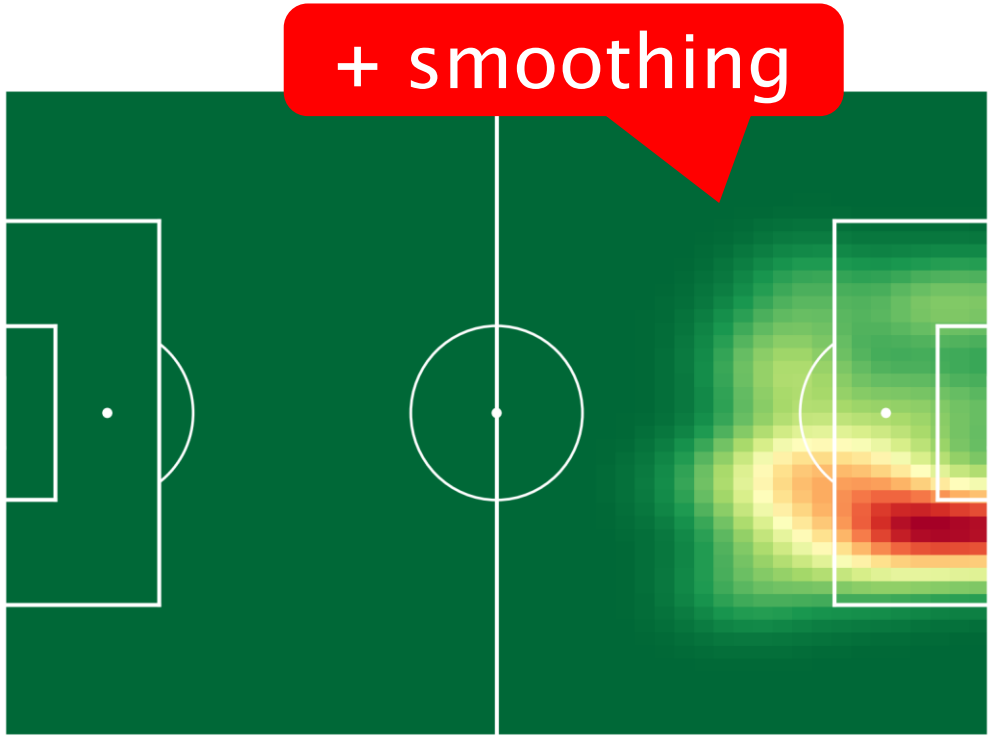
0	0	1
0	1	3
0	0	2

Heatmap (3x3 matrix)

Real example: Shot heatmap of Riyad Mahrez



Scatterplot



Heatmap (50x50 matrix)

Building player vectors in 4 steps

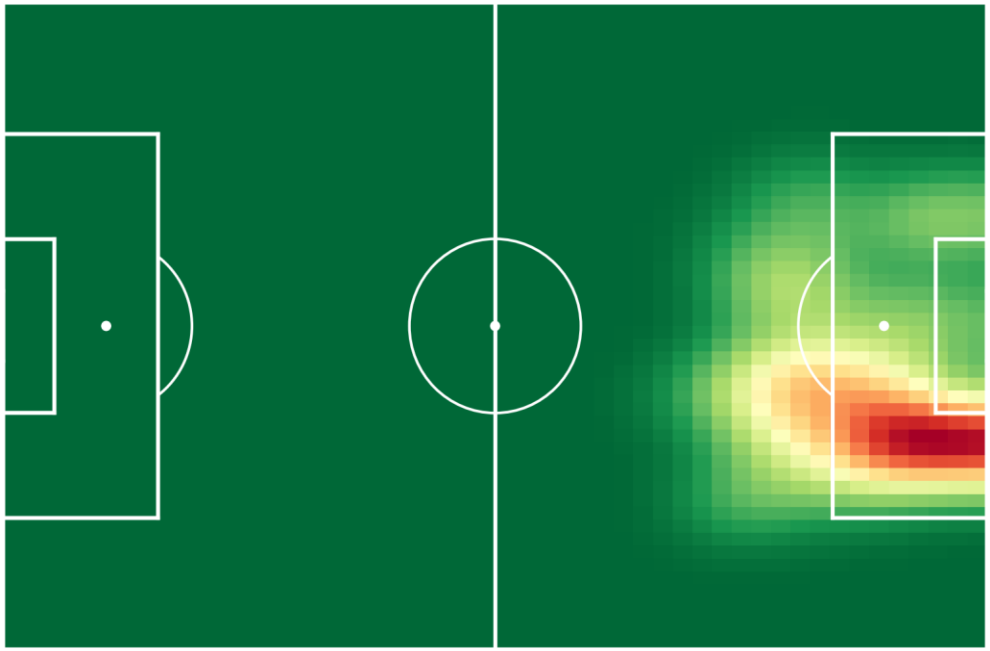
Given: Event stream data of a player p

Do: Characterize player p's playing style in a vector that is human-interpretable and suitable for data analysis

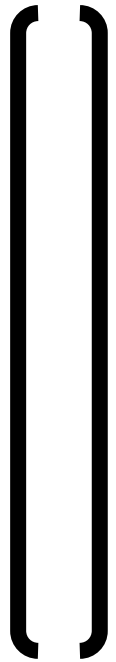
Approach:

1) Select relevant action types
2) Construct heatmaps per player per action type
3) Compress heatmaps to vectors
4)

Flatten shot heatmap to shot vector

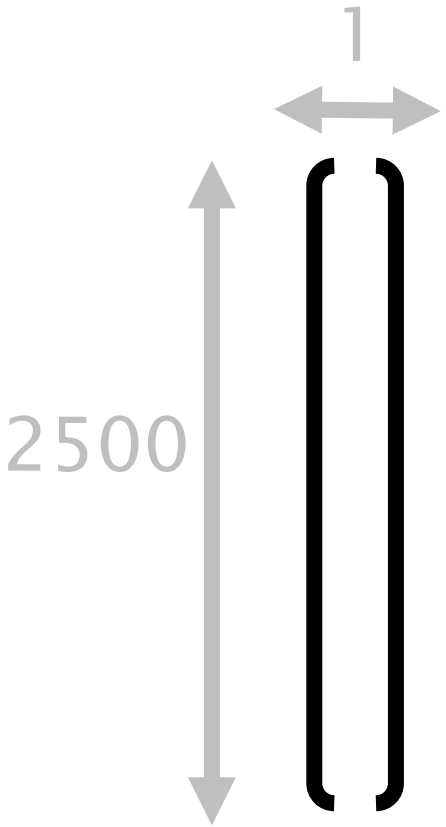


50x50 matrix



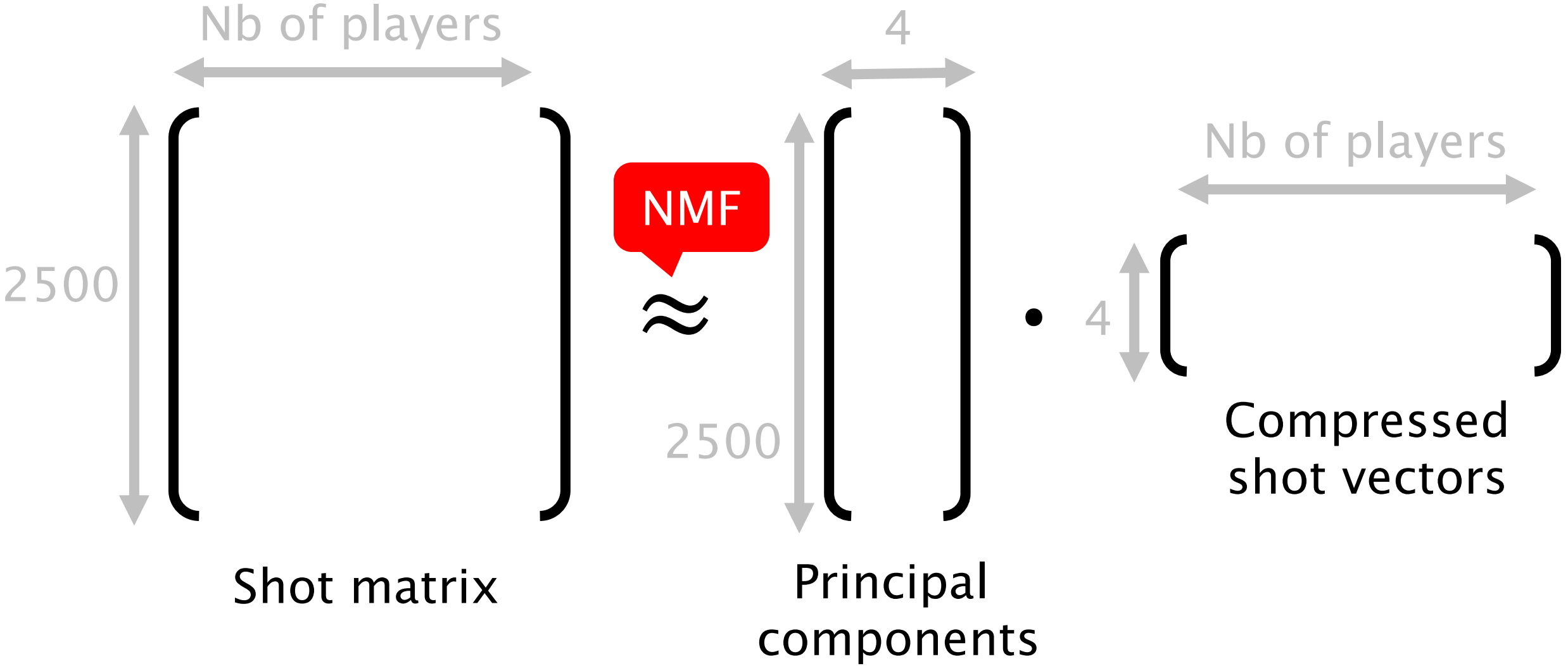
2500x1 vector

Compress shot vectors of all players with Non-negative Matrix Factorization (NMF)

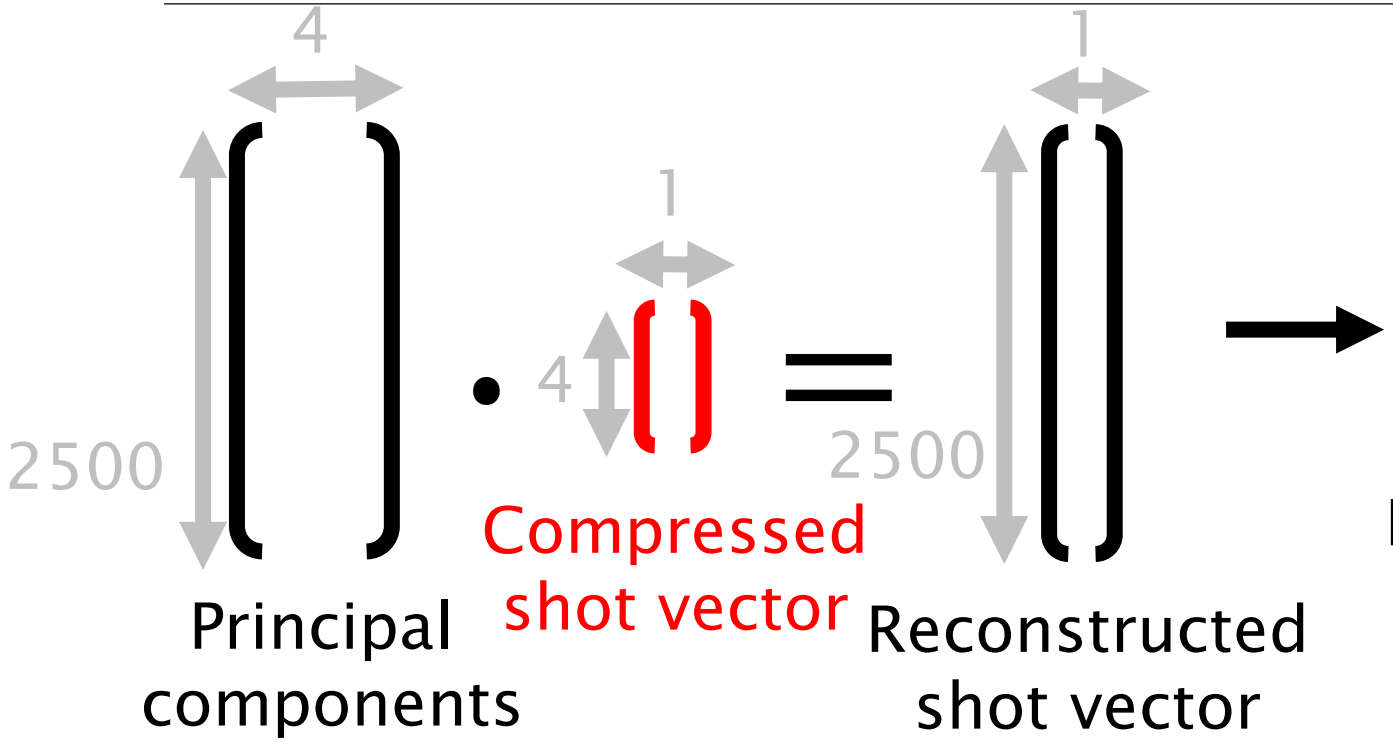


Shot vector

Compress shot vectors of all players with Non-negative Matrix Factorization (NMF)



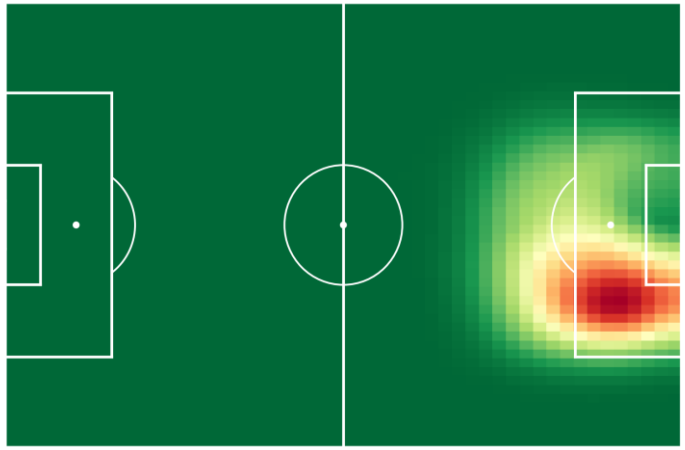
Length-4 vector can reconstruct player shot heatmap



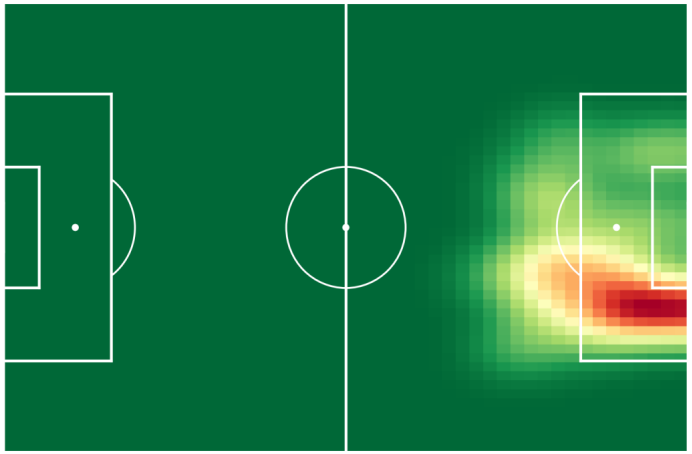
Principal components

Compressed shot vector

Reconstructed shot vector

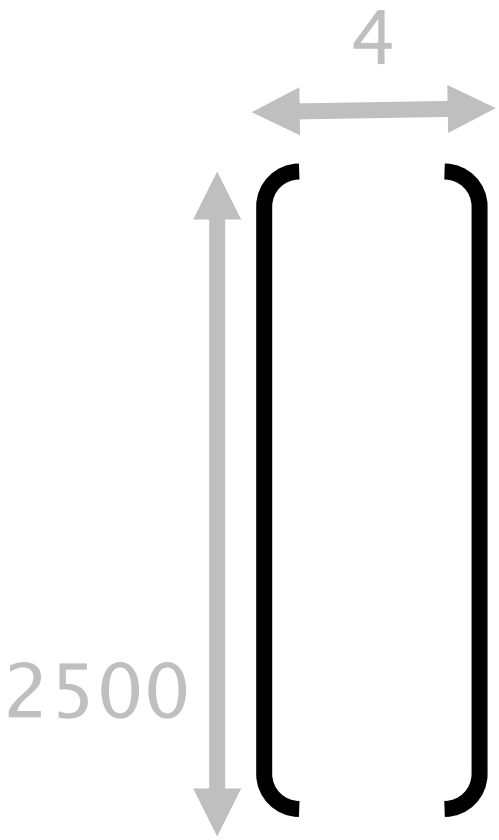


Reconstructed shot heatmap

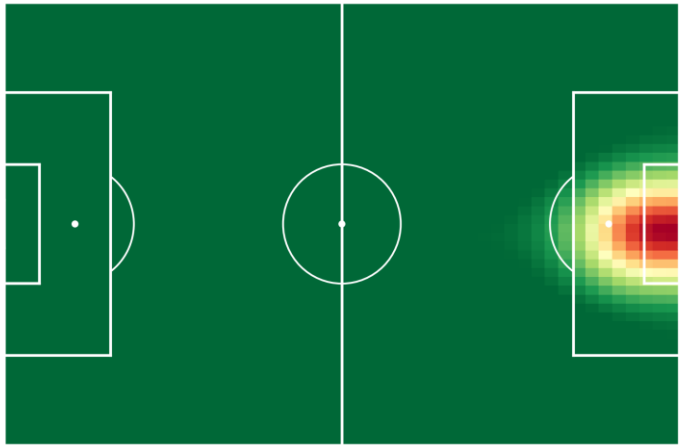


Original shot heatmap

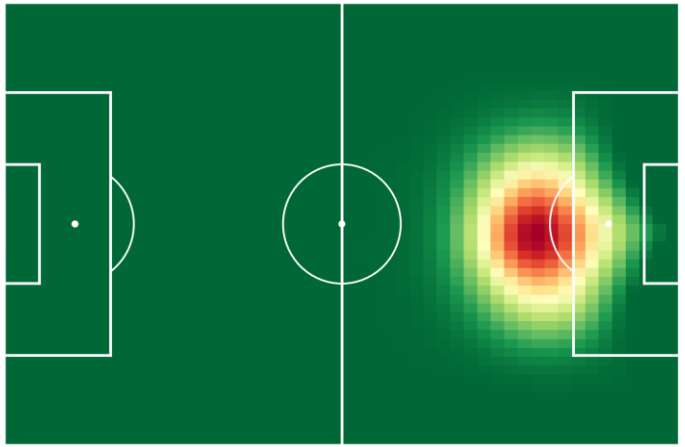
Principal components are archetypical shot types



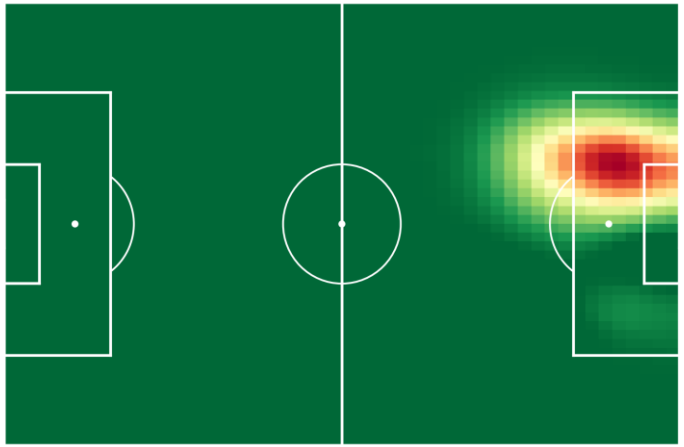
Principal components



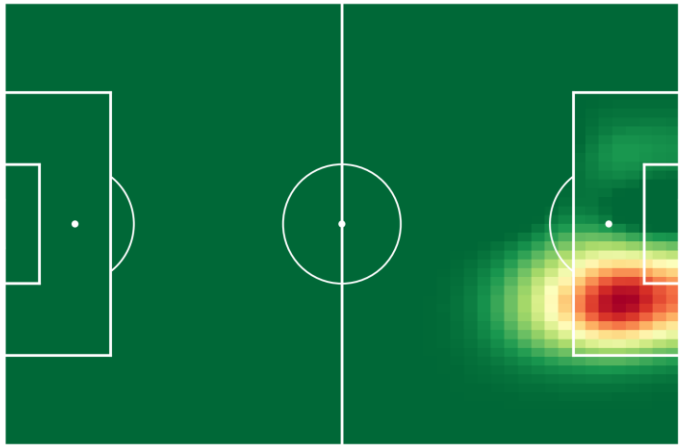
Close



Far

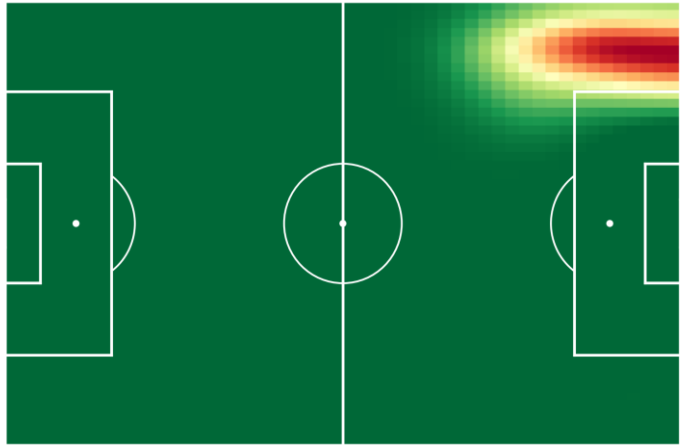


Left

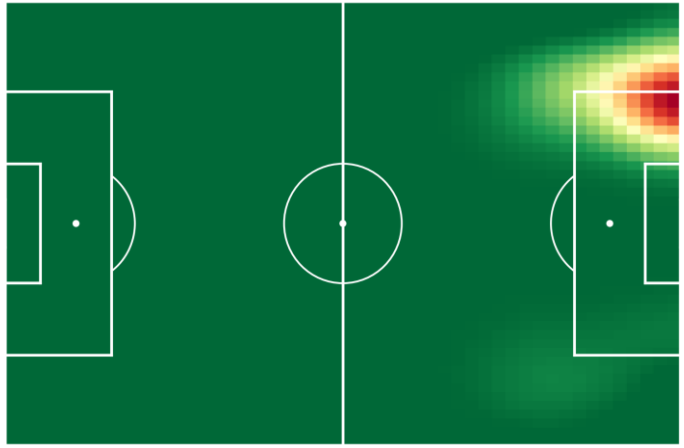


Right

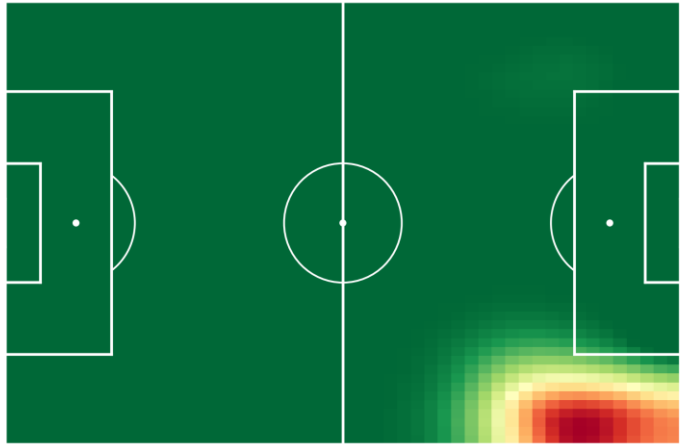
Example: 4 archetypical cross types



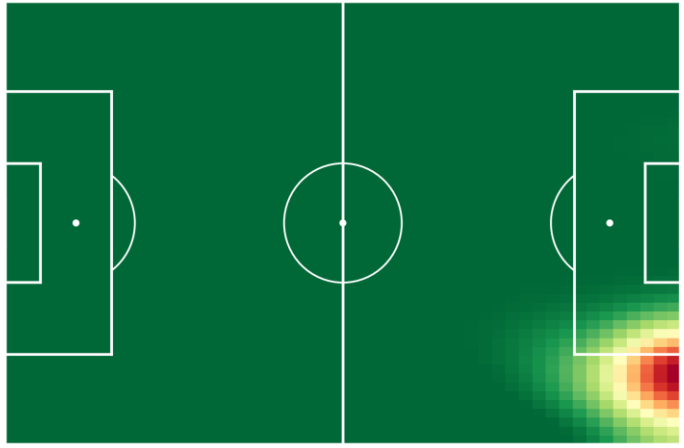
Left flank



Left backline

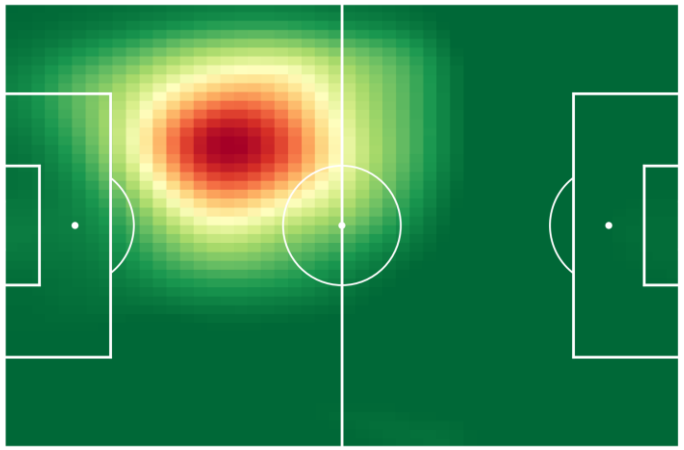


Right flank

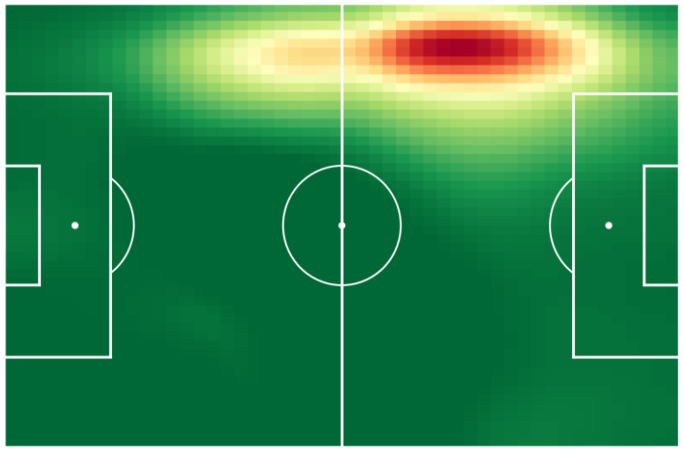


Right backline

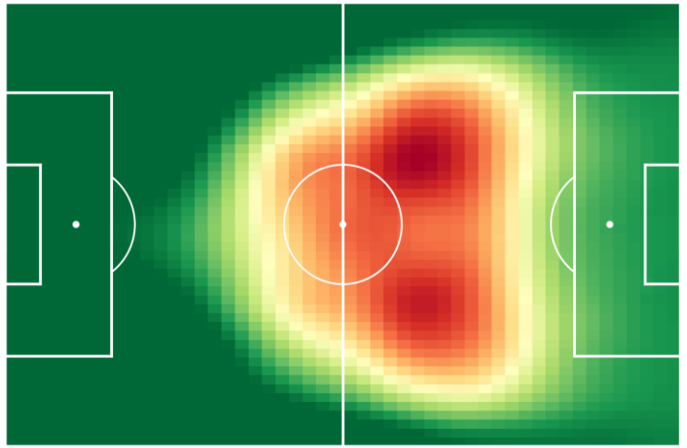
Example: 5 archetypical dribble types



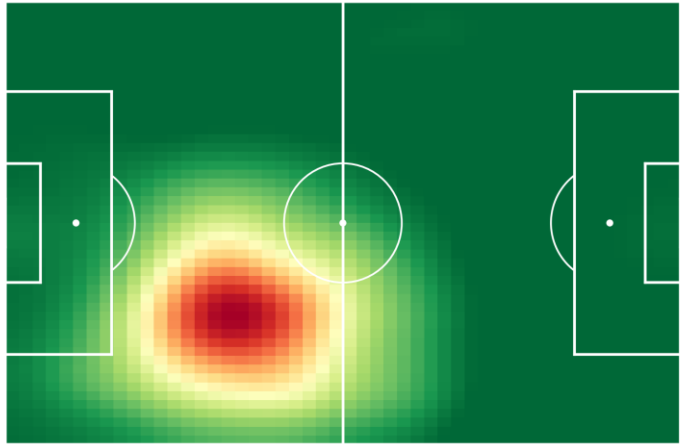
Left back



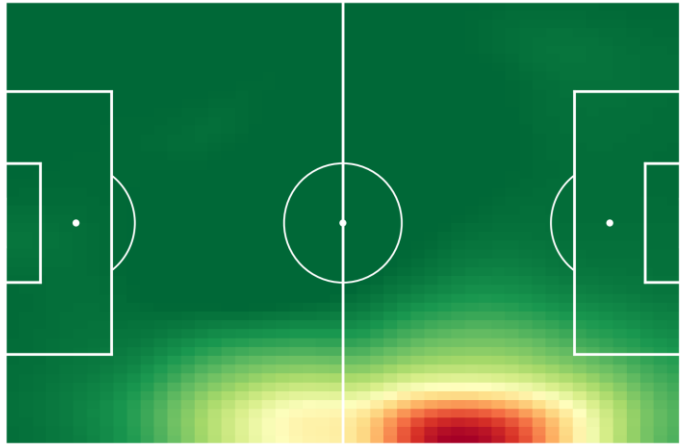
Left flank



Center

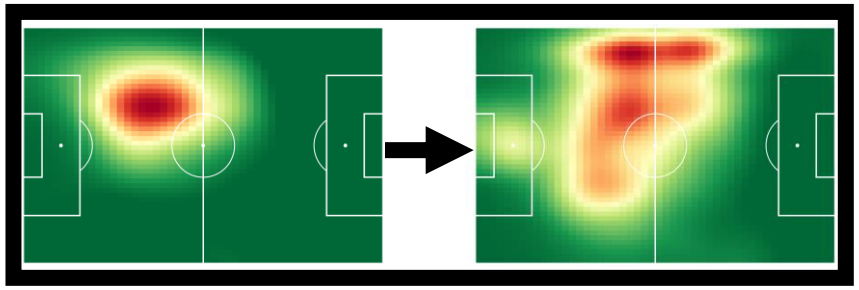


Right back

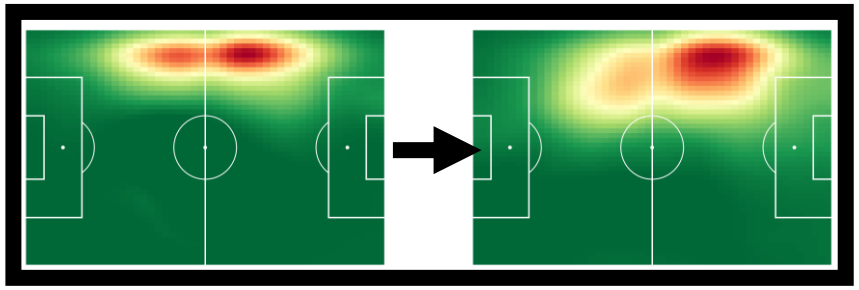


Right flank

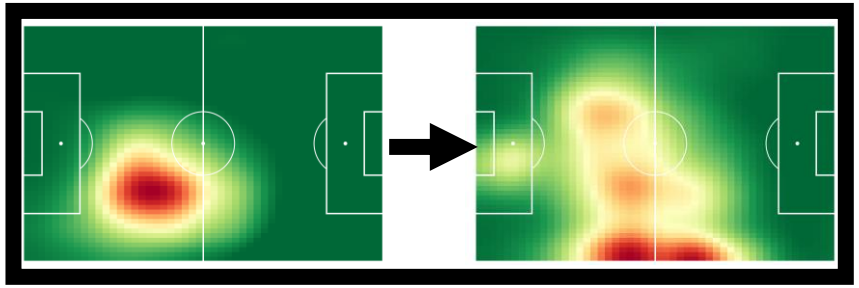
Example: 5 archetypical pass types



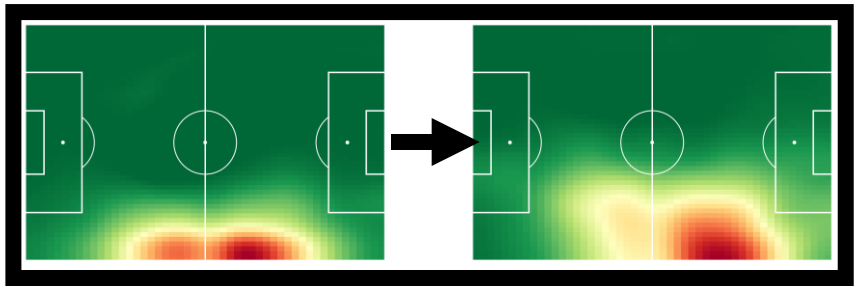
Left back to flank



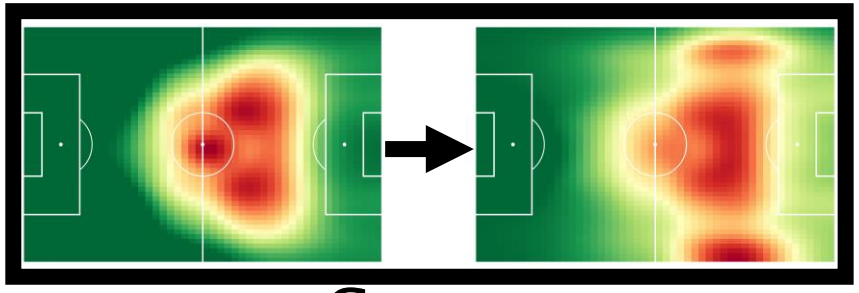
Left flank



Right back to flank



Right flank



Center

Building player vectors in 4 steps

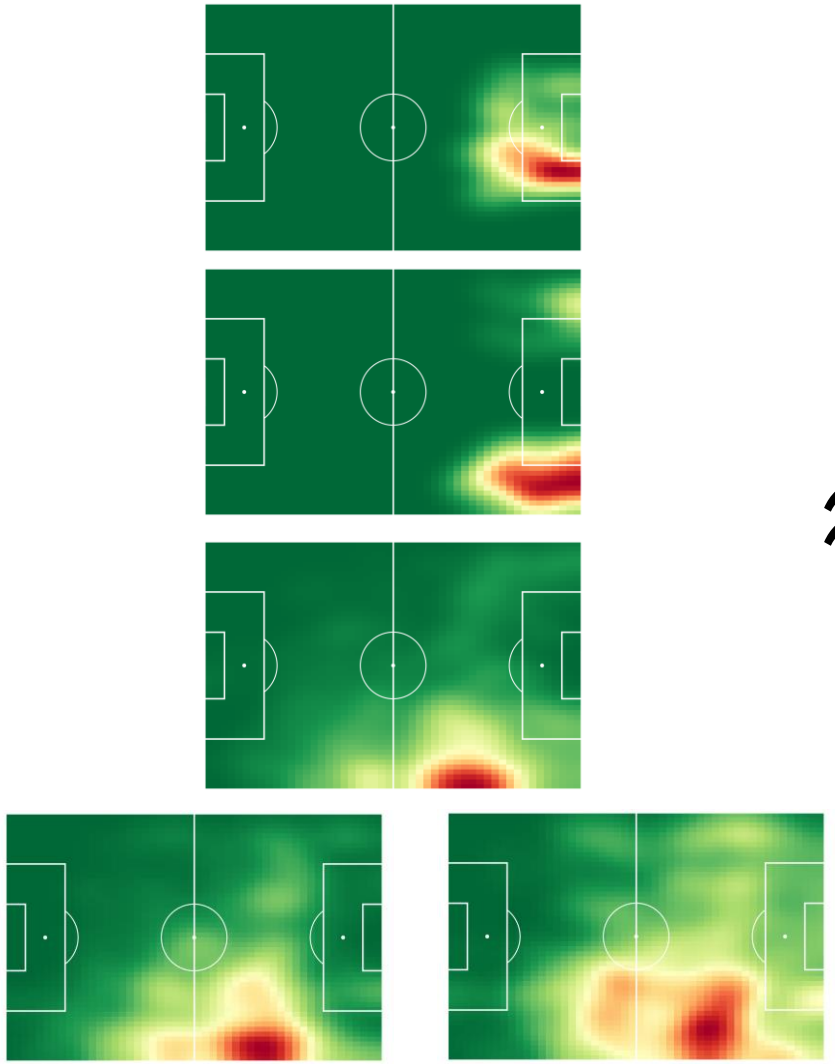
Given: Event stream data of a player p

Do: Characterize player p's playing style in a vector that is human-interpretable and suitable for data analysis

Approach:

1) Select relevant action types
2) Construct heatmaps per player per action type
3) Compress heatmaps to vectors
4) Assemble player vectors

Assemble player vectors



≈

4 Shot components

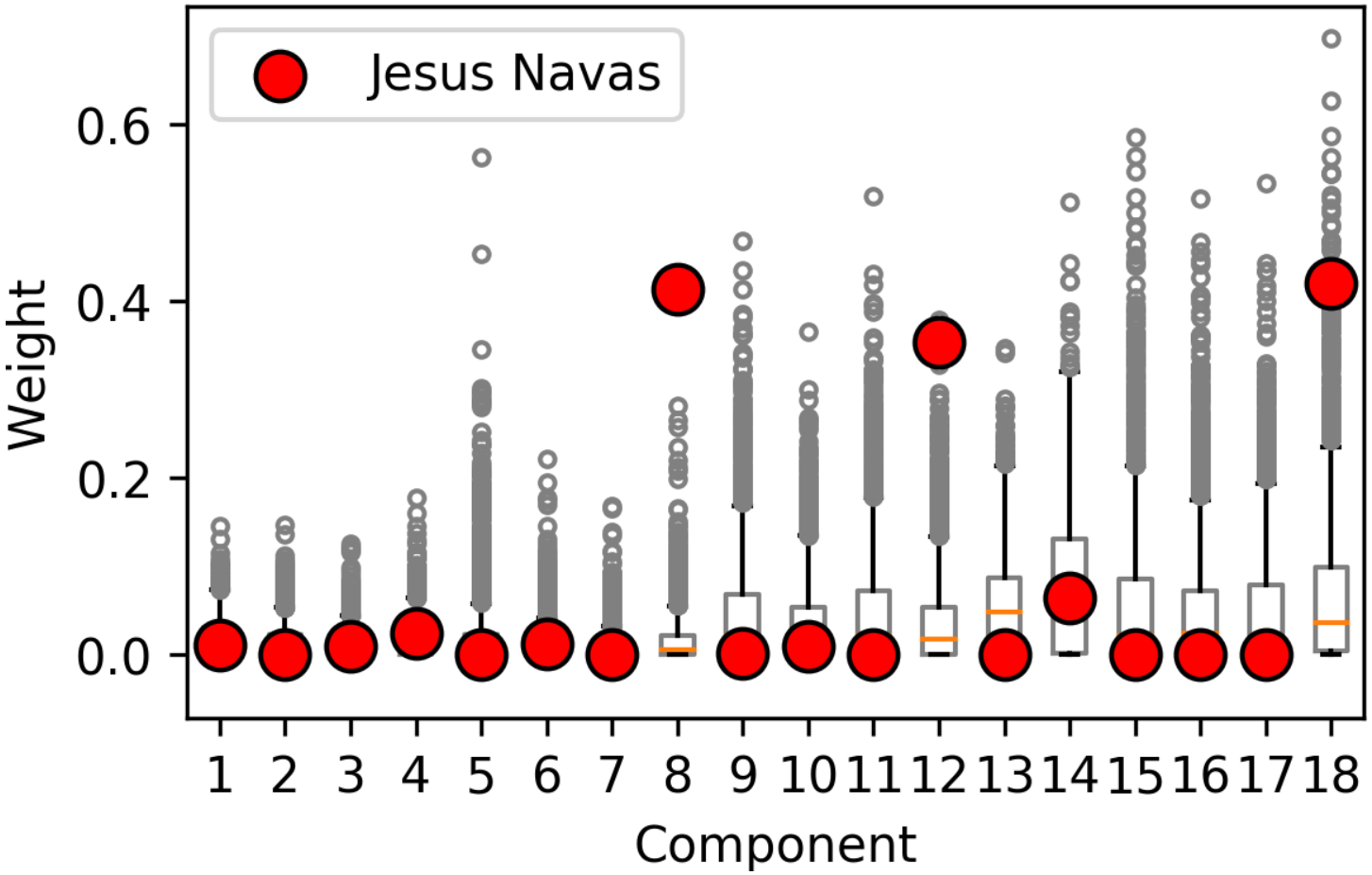
4 Cross components

5 Dribble components

5 Pass components

Example: Player vector of Jesus Navas

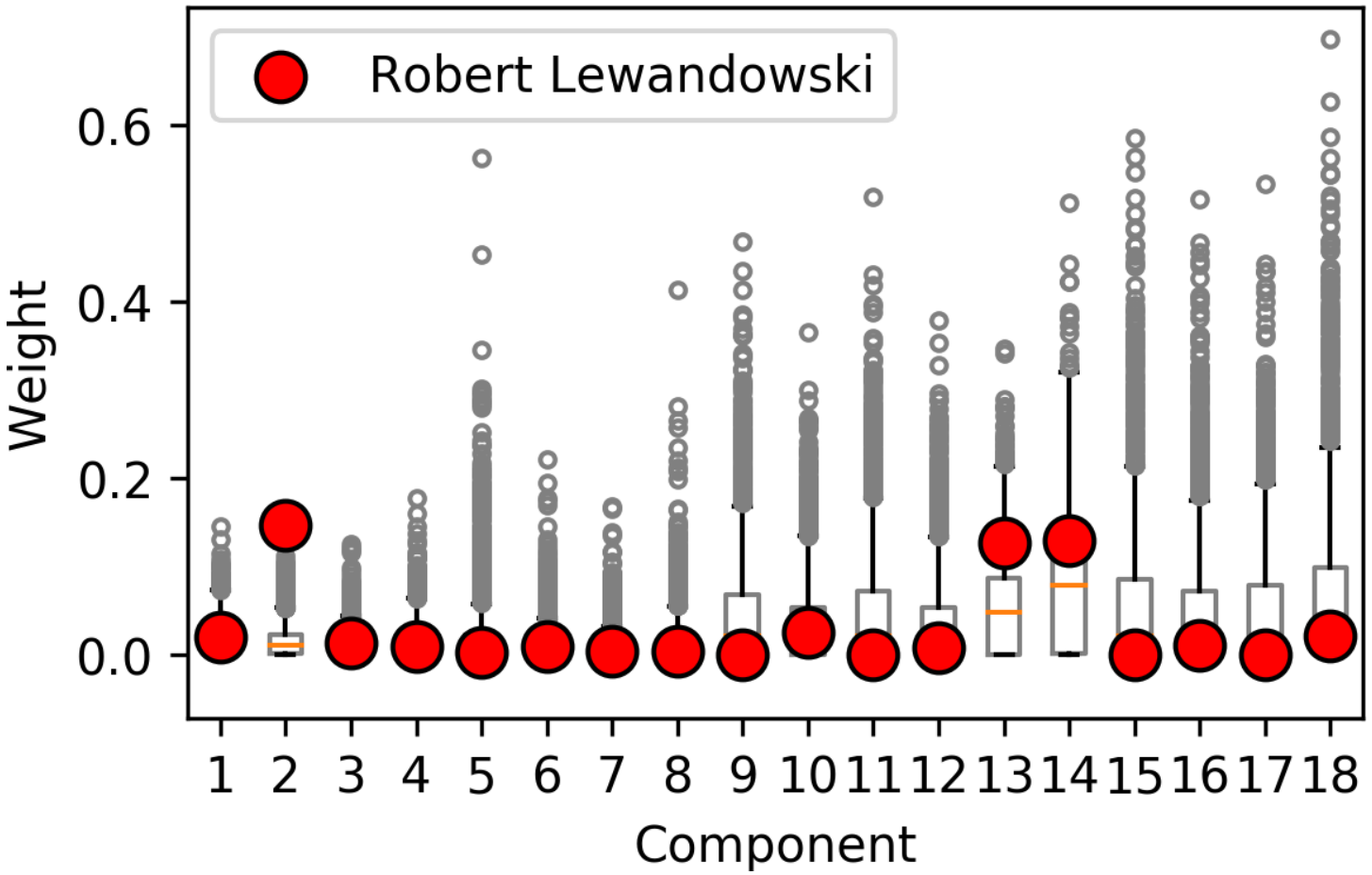
Right winger



High weights for:

- C8: Right backline cross
- C12: Right flank dribble
- C18: Right flank pass

Example: Player vector of Robert Lewandowski

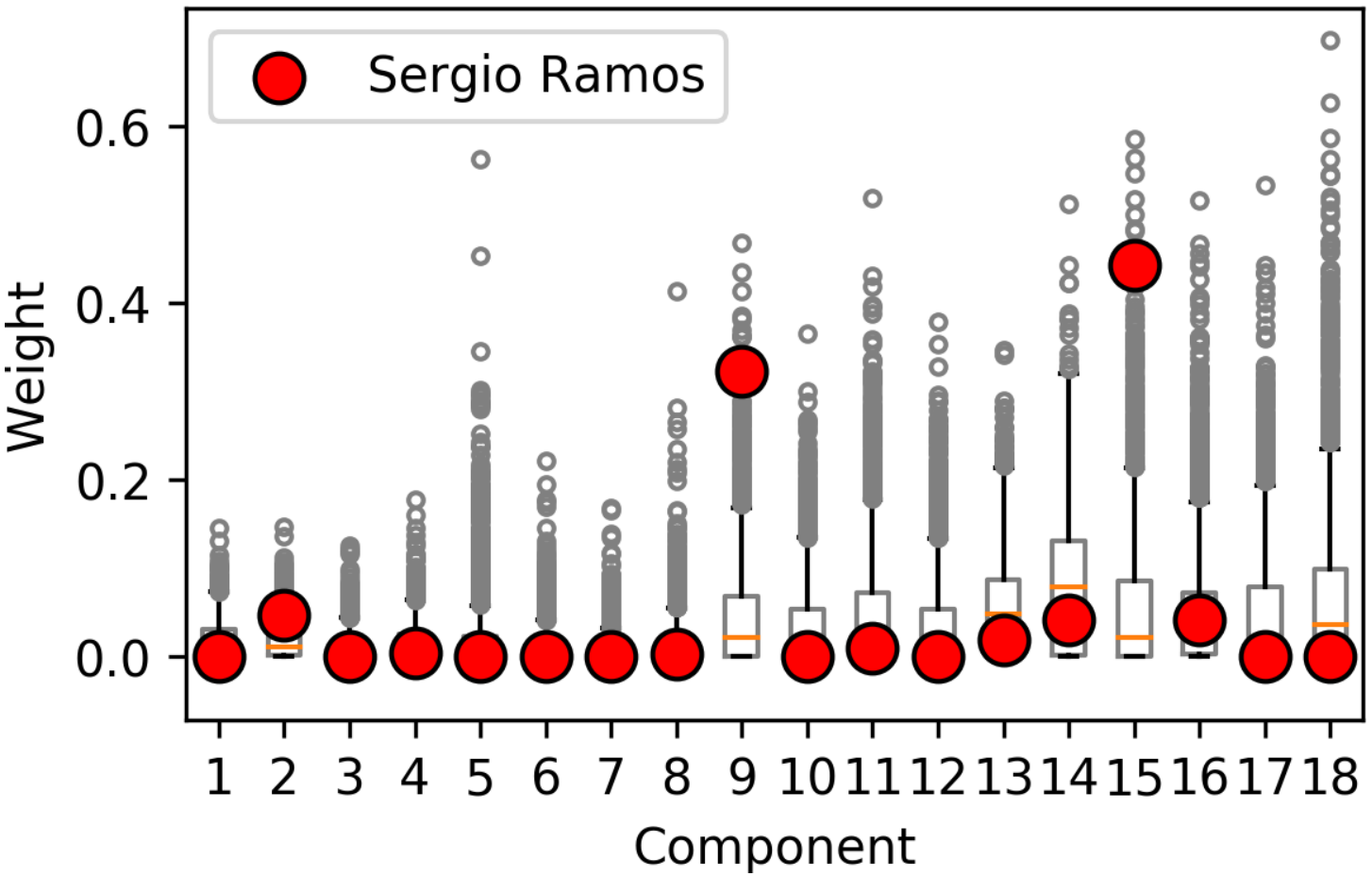


Central striker

High weights for:

- C2: Close shot
- C13: Center dribble
- C14: Center pass

Example: Player vector of Sergio Ramos



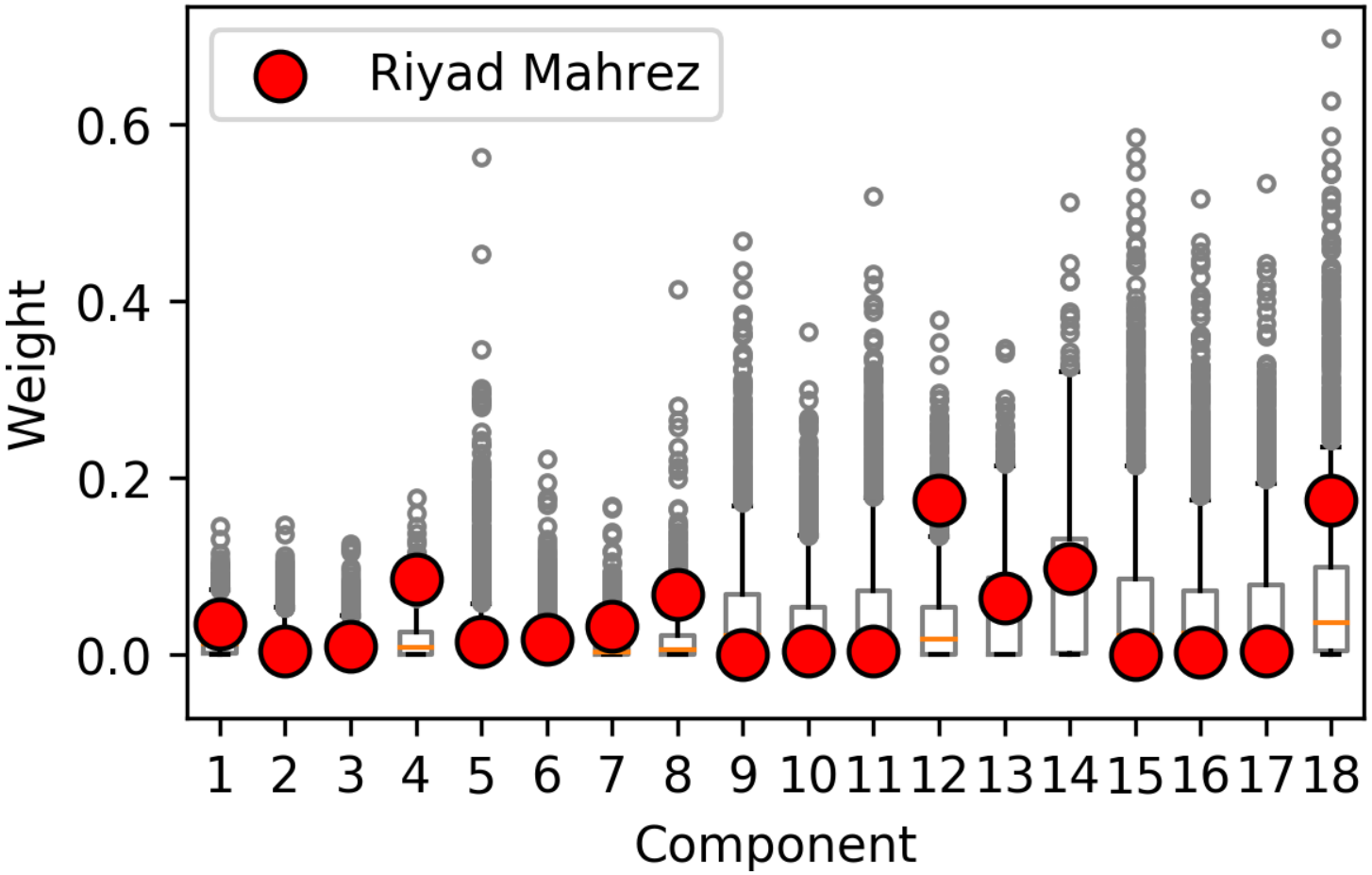
Left-most central defender

High weights for:

- C9: Left back dribble
- C15: Left back to flank pass

Example: Player vector of Riyad Mahrez

Right Winger



High weights for:

- C4: Right shot
- C8: Right backline cross
- C12: Right flank dribble
- C18: Right flank pass

Outline

- Data and Challenges
- Building Player Vectors
- Experiments

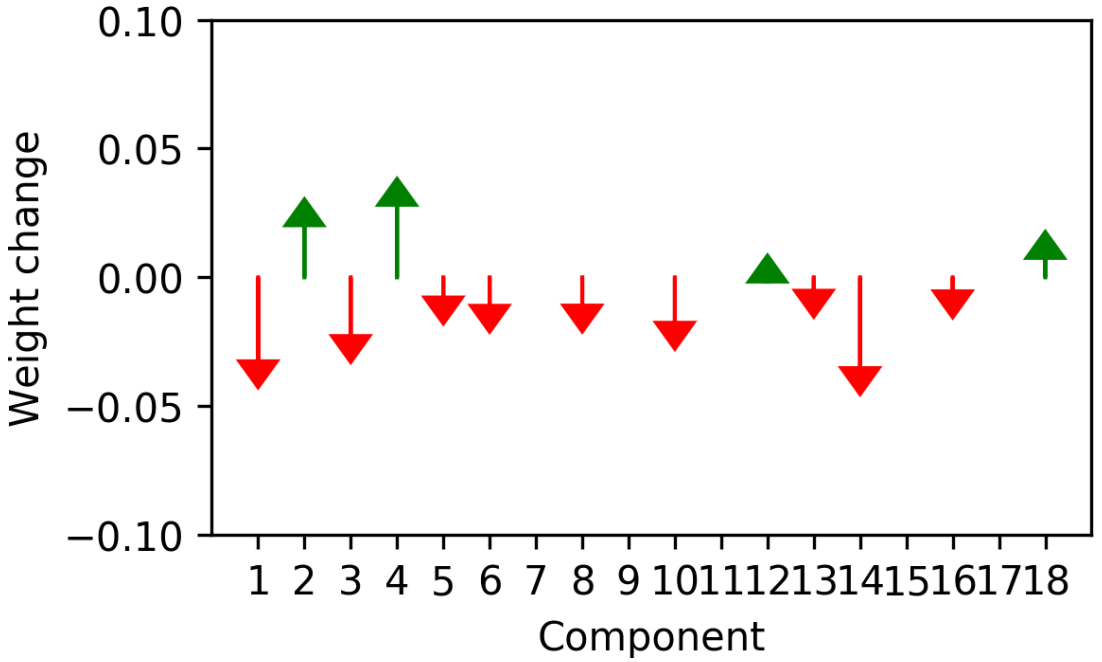
How similar are two players in playing style?

How similar are two players in playing style?
= What is the distance between their player vectors?

Example: Top-5 most similar players to Lionel Messi:

1	Josip Ilicic	Fiorentina
2	Paulo Dybala	Juventus
3	Iago Aspas	Celta Vigo
4	Thomas Muller	Bayern Munchen
5	Rony Lopes	Lille

Ronaldo's evolution from a left-winger (2012/13) to a central striker (2016/17)



C1: Far shot

C3: Left shot

C2: Close shot

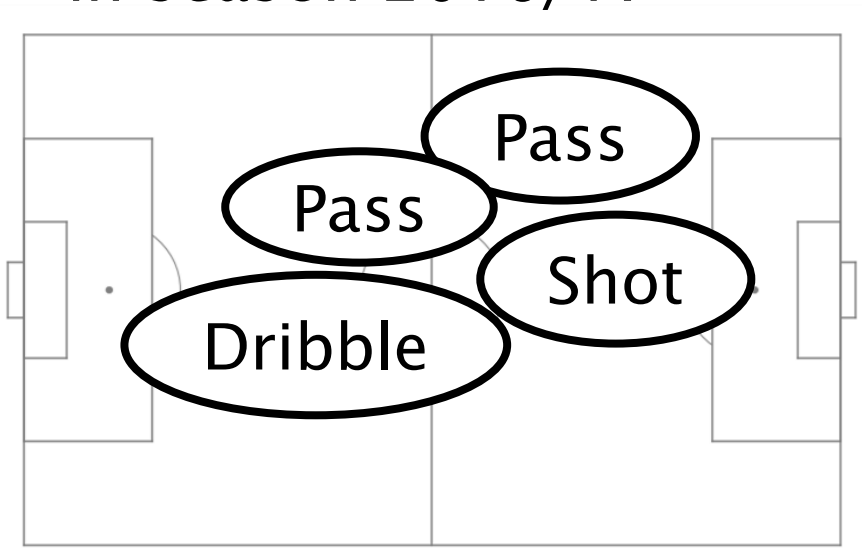
C4: Right shot

Crosses (C5-C8)

How to tune parameters without ground truth on playing style similarity?

Key insight: players are similar to themselves

Eden Hazard's event data in season 2016/17



Most similar players in season 2015/16

1	Alexis Sanchez	Arsenal
2	Alex Iwobi	Arsenal
3	Eden Hazard	Chelsea
4	Muniain	Ath. Bilbao
5	Julian Brandt	Leverkusen

Example: Manhattan distance or Euclidean distance?

Training data: event data of 960 players in season 2015/16
Test data: event data of 960 players in season 2016/17

Ranking quality metrics

Distance function	Top-1	Top-3	Top-5	Top-10	MRR
Manhattan distance	**38.2%**	**49.8%**	**54.9%**	**64.4%**	**0.469**
Euclidean distance	33.0%	47.0%	52.9%	61.8%	0.429

Conclusions

Characterizing playing style from event data is challenging

- No universal definition of playing style
- Handling the spatial component

Player vectors quantify playing style and are

- Human-interpretable
- Suitable for data analysis (e.g., nearest neighbors, clustering, …)