BARÇA SPORTS ANALYTICS SUMMIT

SPORT LOGiQ

BARÇA INNOVATION HUB
Reinventing Sport

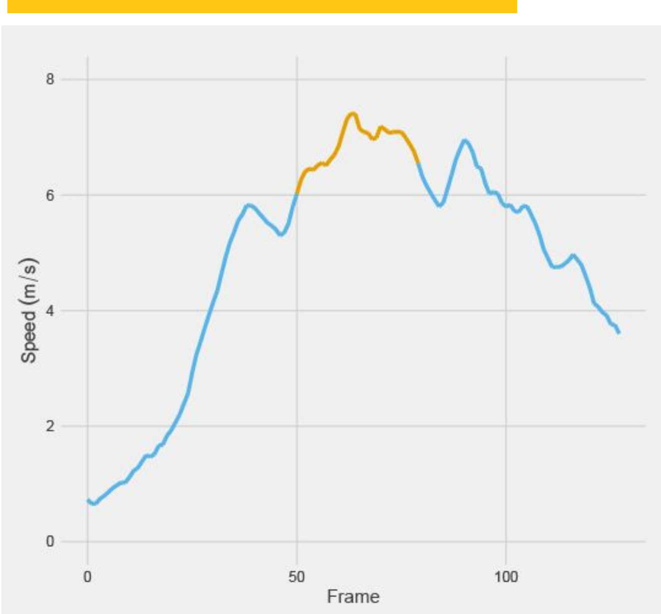
Ready Player Run: Off ball run identification and classification

Sam Gregory / Sportlogiq / sam@sportlogiq.com / @gregorydsam

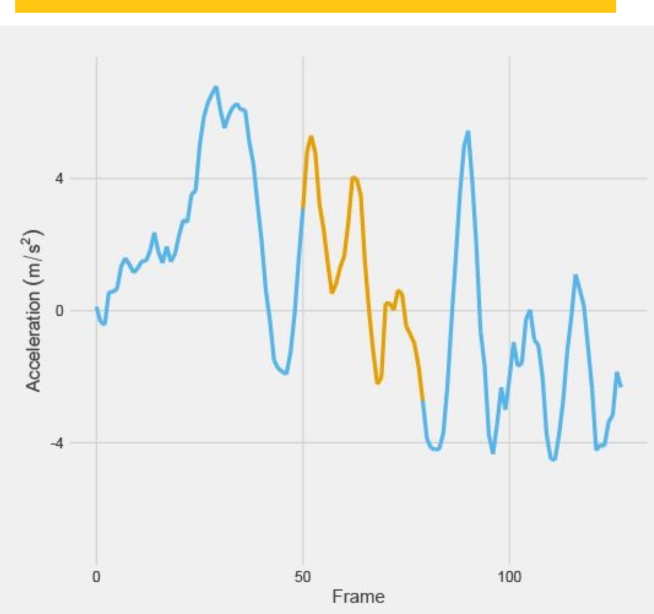
Abstract

One of the major downfalls of tracking data in football is the lack of a common language to describe actions that take place off the ball, particularly patterns of player movement. This approach provides a method for identifying and classifying off-ball in-possession runs into similar groups to allow for more generalisable analysis. The objective is to create a vocabulary of run types that can be used to better describe or analyse specific runs and be queried more easily than raw tracking data. These runs are identified by segmenting the raw tracking data using periods of high player speed and acceleration, then classifying them using a clustering method with functional cluster centres modelled as Bézier curves.

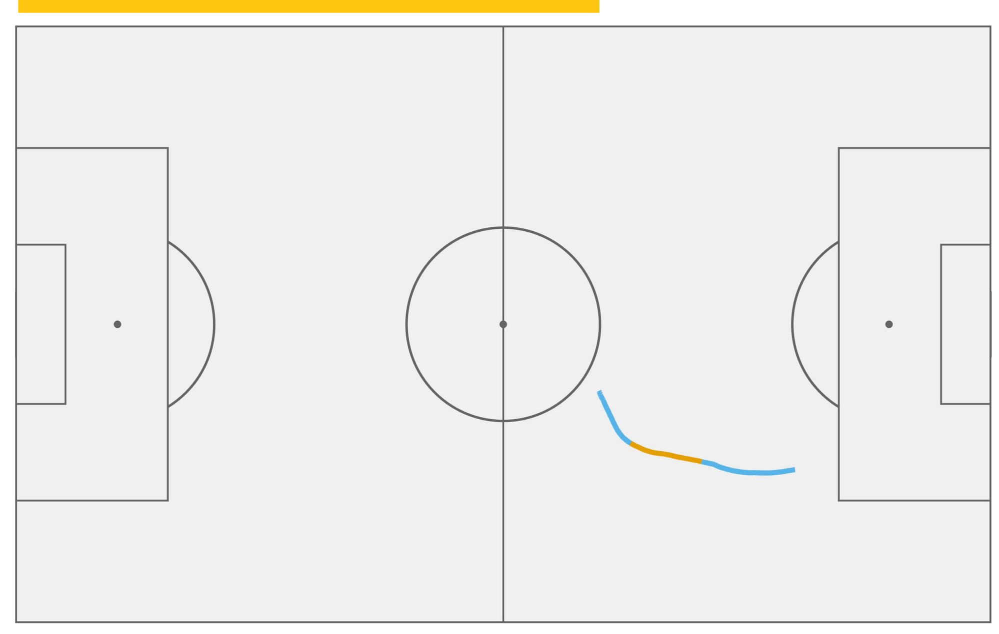
Player Run Speed (+/-2 Seconds)



Player Run Acceleration (+/-2 Seconds)



Player Run Trajectory (+/-2 Seconds)



Run Identification

Run Starts (both conditions met):

- Player speed > 5 m/s, player acceleration
- Player acceleration > 2.5 m/s^2

Run Ends (any condition met):

- Player speed < 4 m/s
- Player acceleration < -2.5 m/s^2
- Team loses possession
- Stoppage in play

Clustering - Bézier Curves

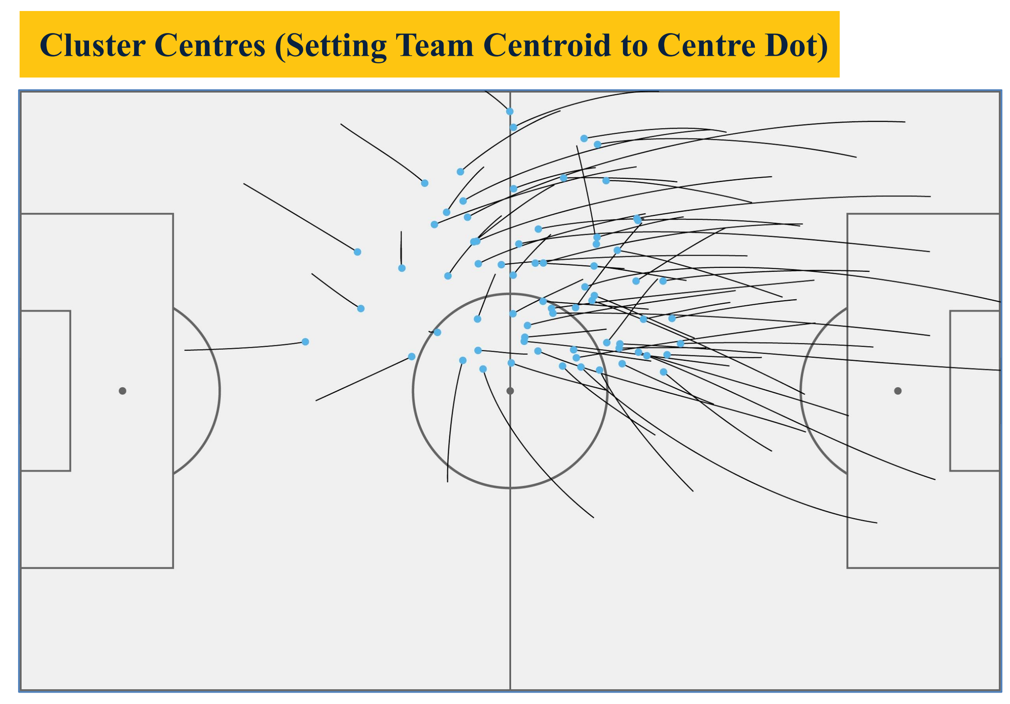
Cluster centres are represented by Bézier curves which provide an ideal functional form for measuring distances between trajectories of arbitrary lengths and cluster centres. These Bézier curves are defined as follows where Θ is the matrix of control points (x,y coordinates) and P is the number of control points.

$$B(t,\theta) = \sum_{p=0}^{P} \theta_p D_p^P(t), t\epsilon[0,1] \qquad D_p^P(t) = \binom{P}{p} t_p (1-t)^{(P-p)}, for\ p = 0,..,P$$

The cluster centres (control points) are determined by solving the following linear least squares problem and iterating until convergence when the average distance from each trajectory (r_i) of length d_i to its cluster centre is below a certain tolerance level.

$$r_i = X_i\theta + \epsilon_i \qquad X_i = [D_p^P(t_{i1}), \ldots D_p^P(t_{id_i})]$$

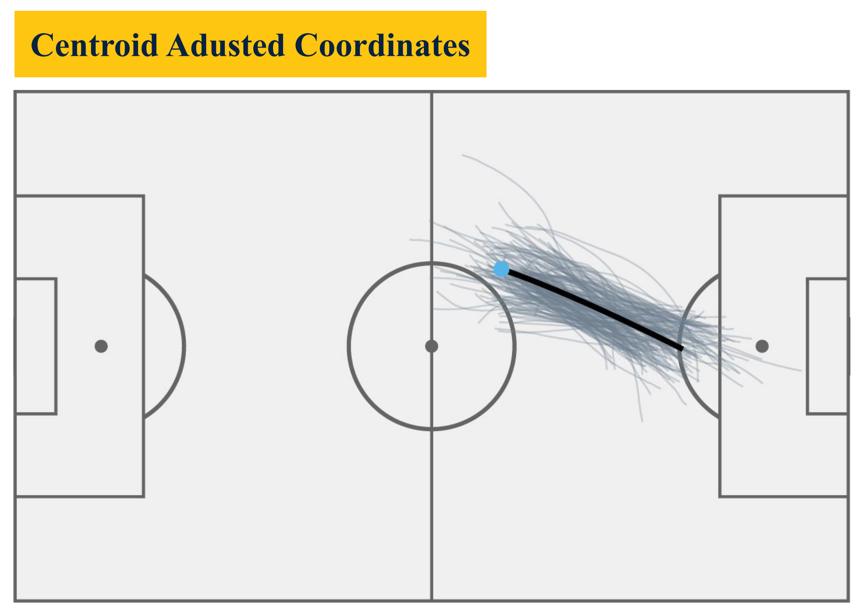
Cluster Centres (Setting Team Centroid to Centre Dot)



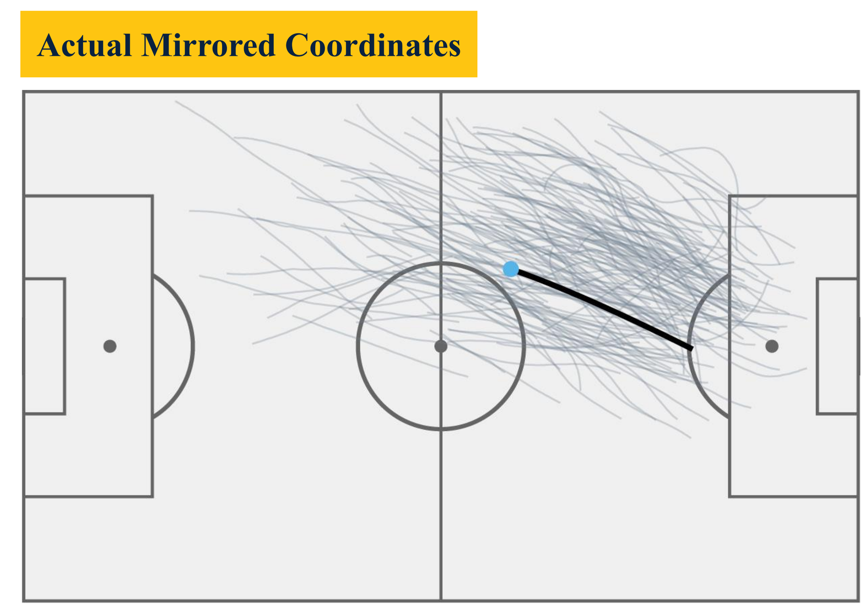
Coordinate Adjustment

The clustering is done on adjusted coordinates where all runs are in coordinates relative to the team centroid rather than the centre of the pitch (also all mirrored to originate from the left hand side of the pitch). This helps group runs of similar intent that occur in different parts of the pitch. Consider all of the runs grouped in one cluster mapped in both adjusted and actual coordinates.

Centroid Adusted Coordinates



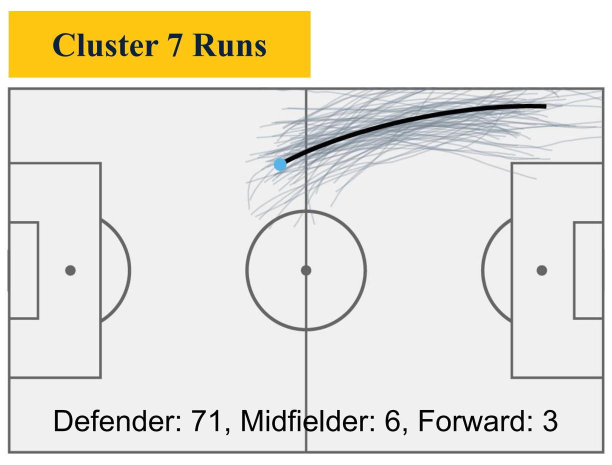
Actual Mirrored Coordinates



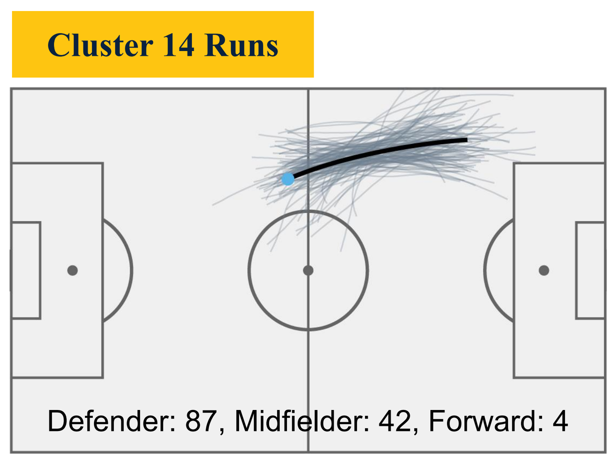
Overlapping Fullback Runs

A grouping of the most common run types made by fullbacks highlights different styles of “overlapping” runs.

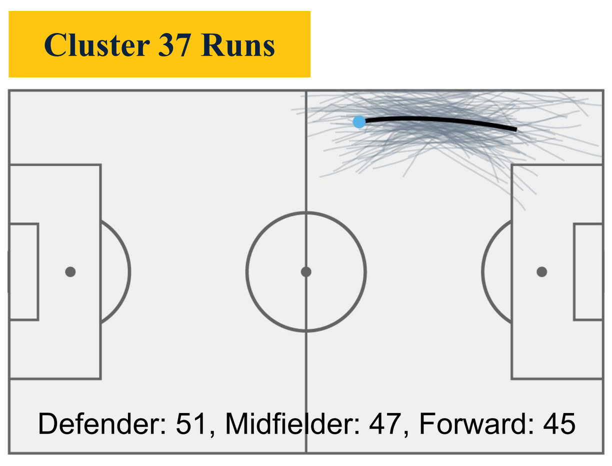
Cluster 7 Runs



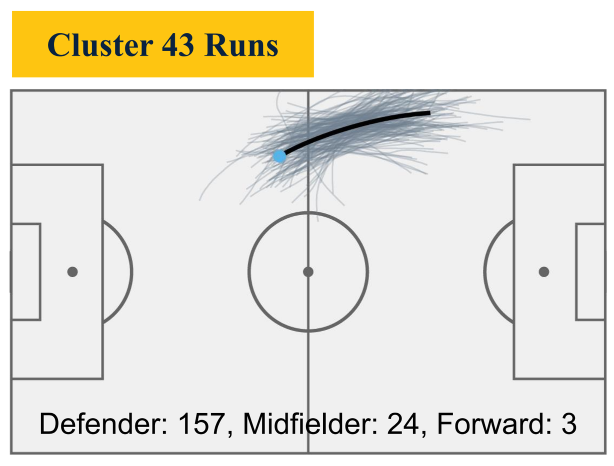
Cluster 14 Runs



Cluster 37 Runs



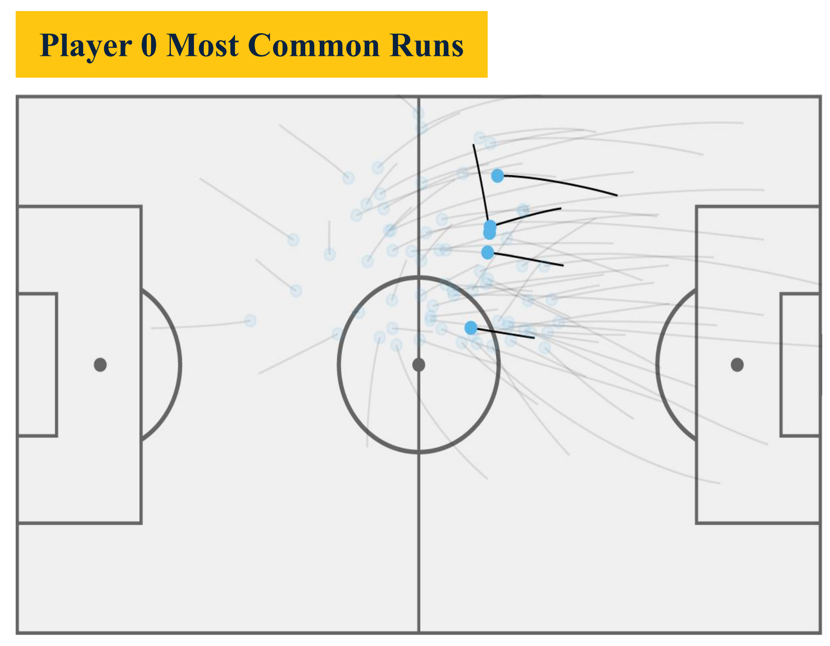
Cluster 43 Runs



Player Archetyping - Most Common Runs

By looking at a player’s most common off-ball in-possession runs we can learn a lot about that player’s tendencies. Looking at these two anonymised players’ most common run types we can clearly identify that player 0 is an attacking player who gets wide to stretch the field and cut in while player 1 is a full back who makes lots of wide overlapping-style runs from positions roughly level with the team centroid.

Player 0 Most Common Runs



Player 1 Most Common Runs