

Football Pass Analysis

Using KMeans Clustering - Connie Teo

TABLE OF CONTENTS

01 Problem Statement

02 Data Sources & Cleaning

03 Team Analysis

Pre modelling EDA
K Means & Cluster Findings

04 Player(s) Analysis

Pre modelling EDA
K Means & Cluster Findings

05 Considerations & Recommendations

01 Problem Statement

I represent the Football Association of Singapore, who are looking to invest actively in womens football locally.

With the recent win in Euros 2022, the England women's football team have improved rapidly, in the last few years.

I've selected this team to study their improvement points in order to provide more insights to our local national team.

Their passing statistics are highlighted as a key point to their success, and this presentation will cover my analysis of their passes using KMeans clustering.

02

Data Source & Cleaning

Statsbomb Package

Data Source is from



Statsbomb is a UK-based football analytics and data visualisation company introducing common data analytics practices seen in business and tech to the world of football analytics.

They provide free event data (limited) to the public for analysis purposes, and invite research papers every year to discover new ways of interpreting data.

Event data captures all of the action that the players perform with the ball, and includes passes, dribbles, crosses, interceptions, tackles, shots, etc.

We will be using the passing data.

02

Data Source

Feature Selection & Data Cleaning



Dropped missing values before cleaning the below passing data

Features used	Explanation	Data Cleaning & Feature Engineering
period	first / second half, extra time. Statsbomb flips the data for second half hence no need to differentiate	nil
location	location of starting pass, start_x & start_y	in a form of [x,y], to loop over the rows and separate x & y to different columns. Most important feature for plotting of cluster
pass_end location	location of end pass, end_x & end_y	
pass_outcome	Complete or incomplete. incomplete includes offset or out of bounds, etc.	categorical data, get_dummies
pass_angle	+ve indicates clockwise, -ve indicates anti-clockwise	numerical data
pass_height	ground, low or high pass	categorical data, get_dummies
pass length	in yards.	numerical data

02 Feature Selection & Data Cleaning

03

Team Analysis

this part of the journey is a
minefield

03

Team Analysis

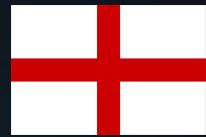
Pre modelling EDA

I've shortlisted 2 tournaments, World Cup 2019, and Euros 2022, to compare England's different styles of play in the 2 events

For reference, England

- Lost to USA 1-2 in the semis in World Cup 2019, and
- Won the Euros in 2022 by defeating Germany 2-1

England played 6 matches in the World Cup 2019 and 5 matches in Euros 2022

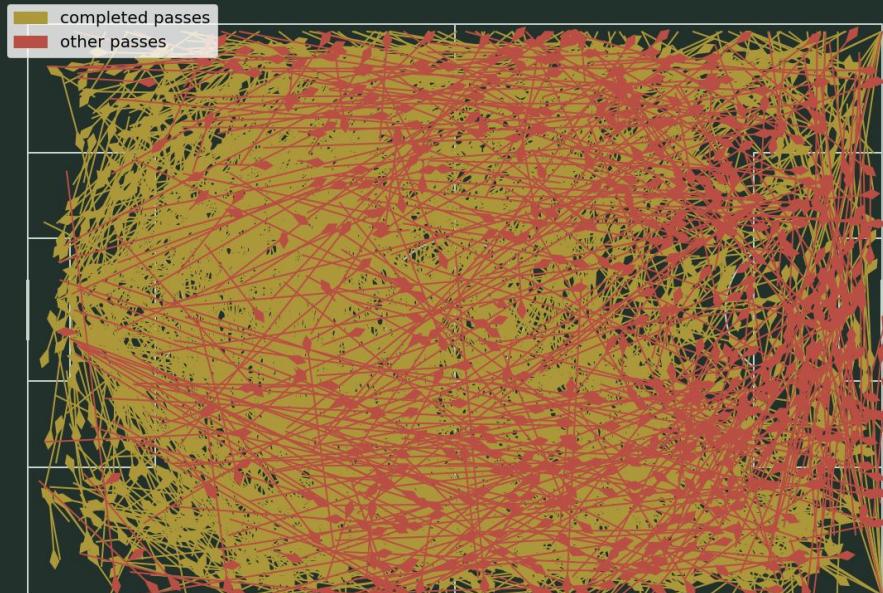


World Cup 2019

Overall team passes

3,808 total passes, **84%** completed

Passes

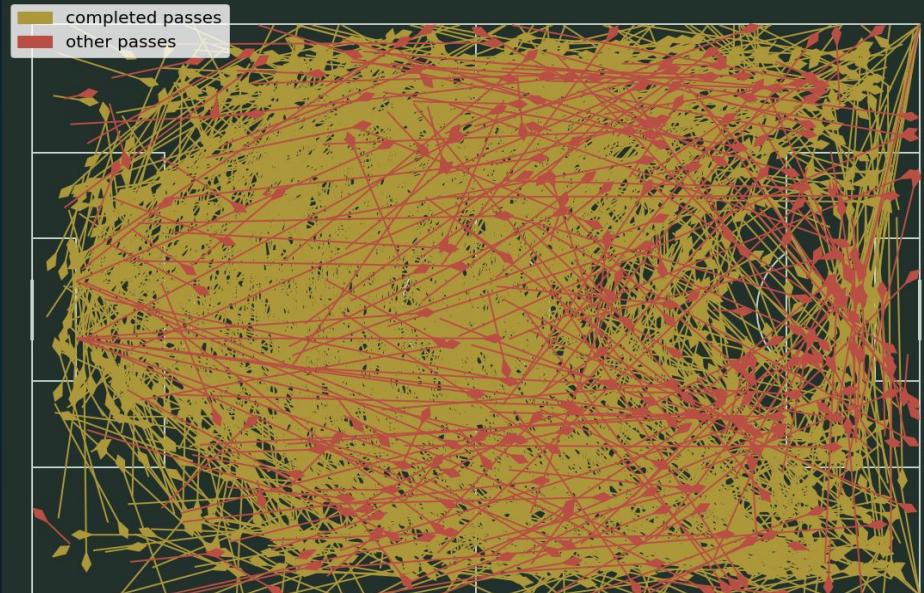


Euro 2022

Overall team passes

2,969 total passes, **91%** completed

Passes



03
Team Analysis

K Means & Cluster
Findings

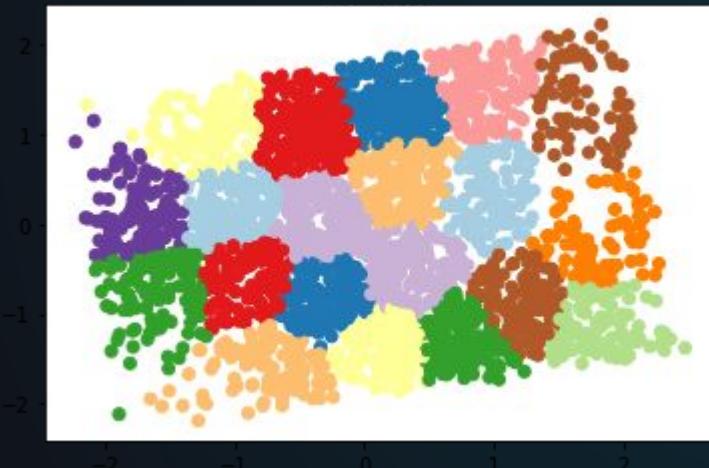
Which method to use

Kmeans with PCA

PCA n_components = 2 to reduce dimensions

StandardScaler
n_clusters = 20

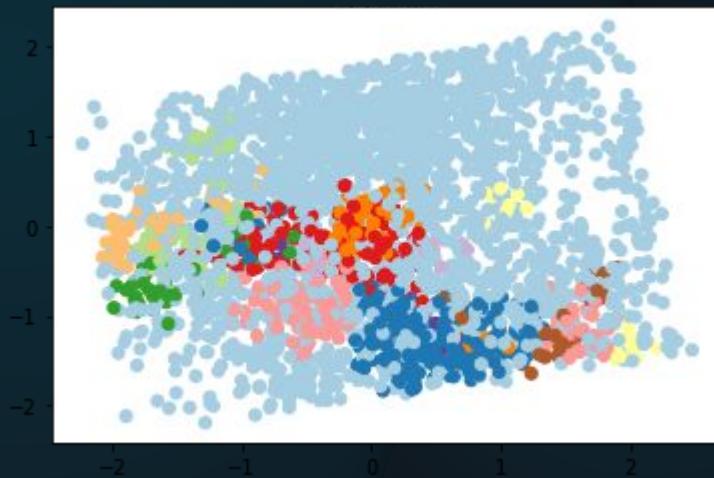
K-means



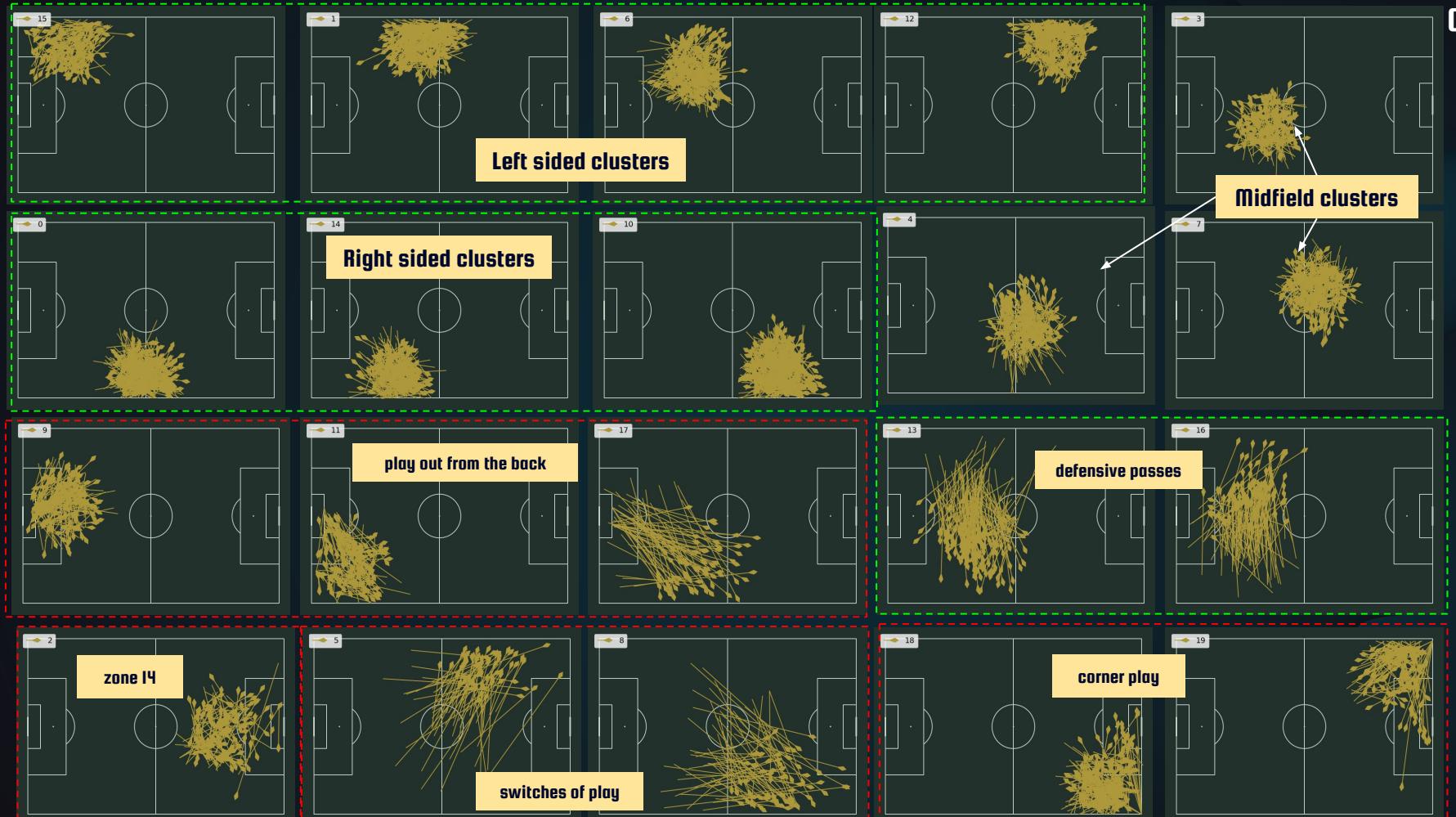
DBScan

StandardScaler
(eps=10, min_samples=10)

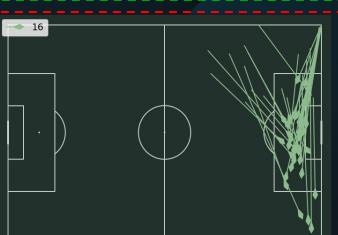
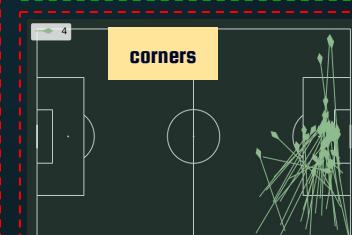
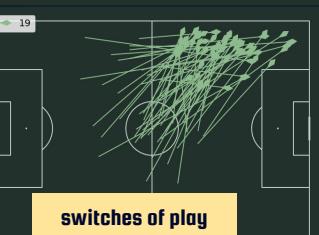
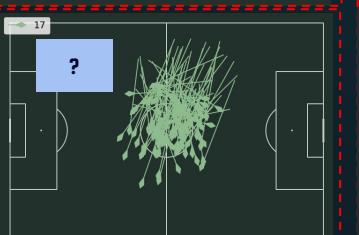
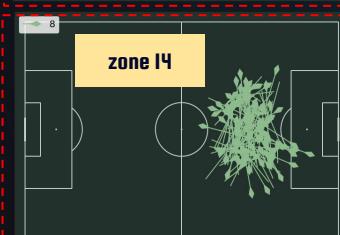
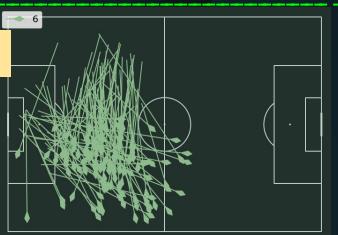
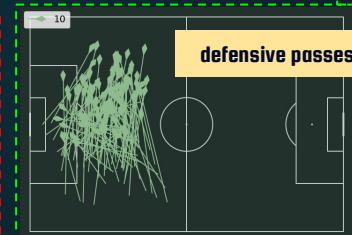
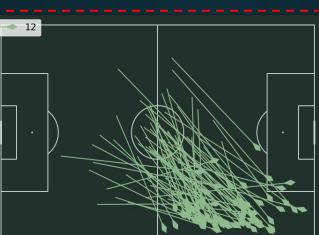
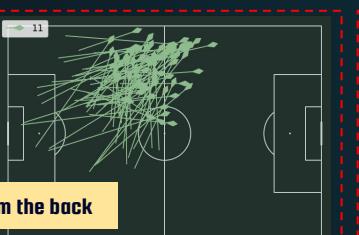
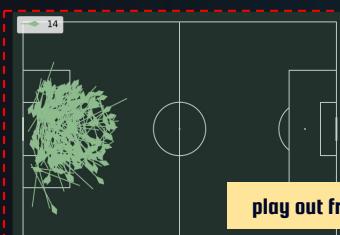
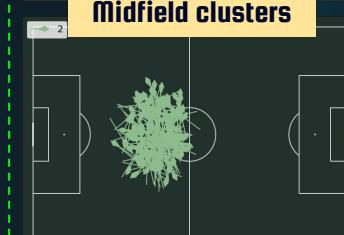
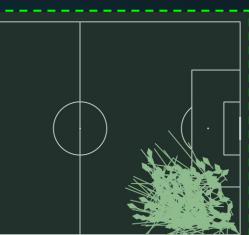
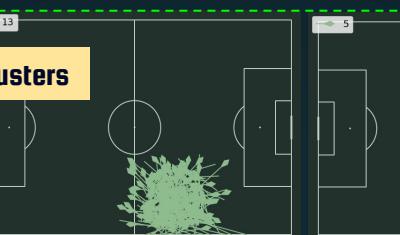
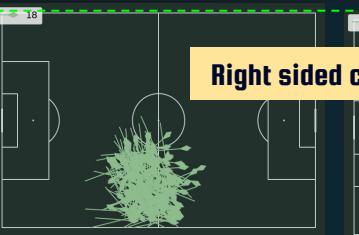
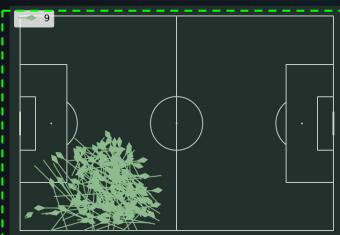
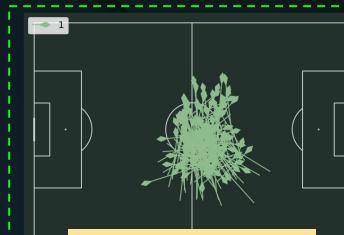
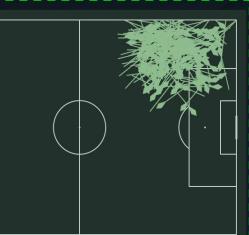
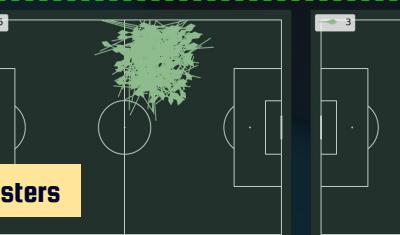
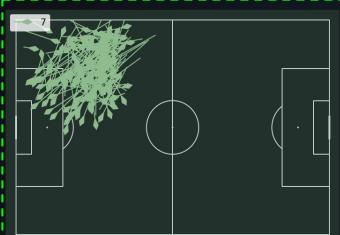
DBScan



Clusters in World Cup 19

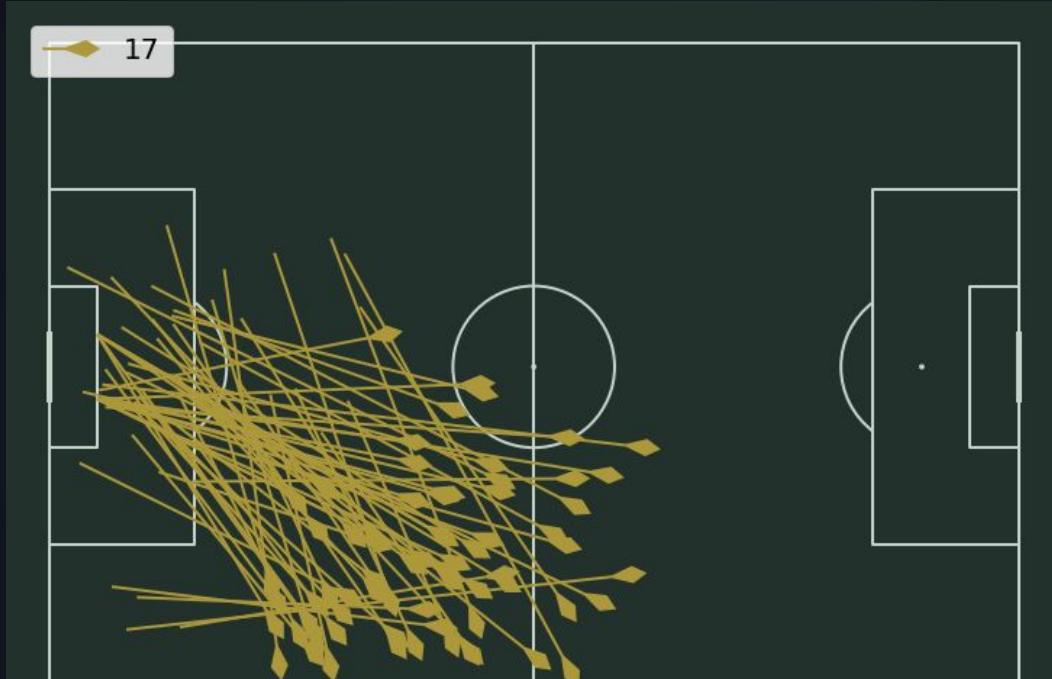


Clusters in Euro 22



switches of play

World Cup 2019



	pass_height_Ground Pass	pass_height_High Pass	pass_height_Low Pass
count	70.000000	70.000000	70.000000
mean	0.185714	0.742857	0.071429

Passes made from the back long & high passes

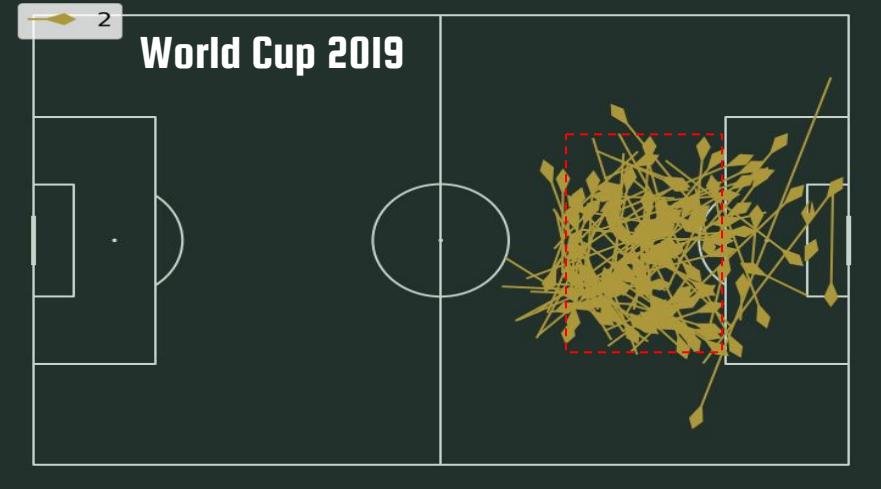
70 passes in cluster

Keeper tends to favour the right for long and high balls

Recipient of these passes are the right full back (Lucy Bronze) + tallest midfielder (Jill Scott)

These passes are not seen in euro 22 as the keeper plays short passes out from the back

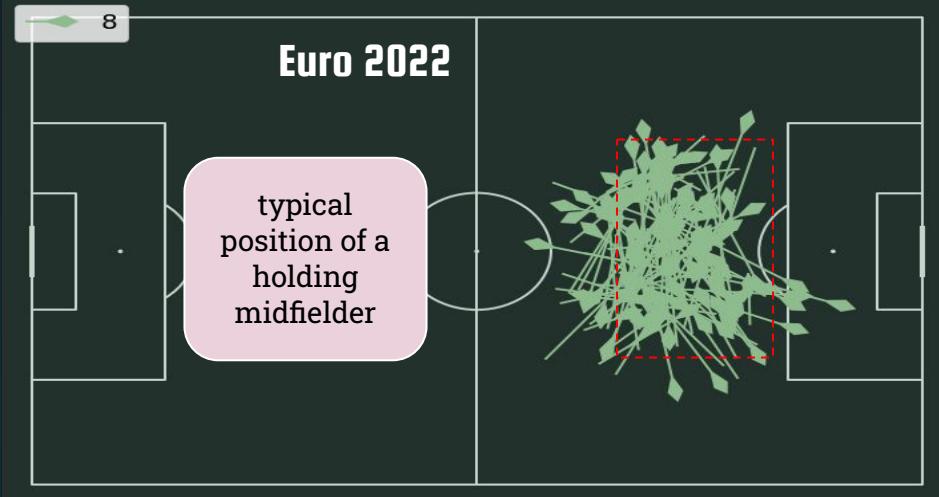
Passes made in zone I4



137 passes in cluster

Key passer and recipient = Fran Kirby (23% of passes). Fran was the playmaker in the team

More through passes reaching the goalmouth



115 passes

A change in tactical play is seen. Keira Walsh is the player most active in this area (25% of passes). She plays holding midfielder for England.

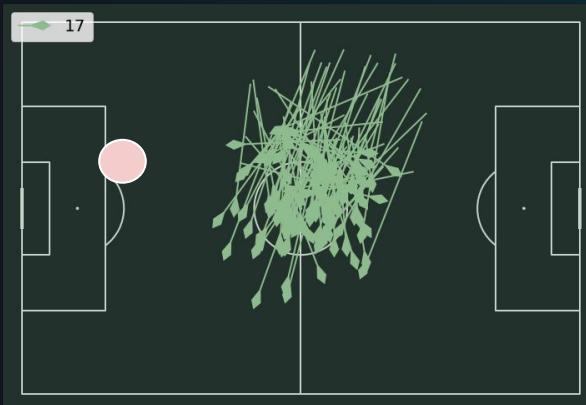
Fran Kirby still plays the no. 10 role but it seems more subdued in this tournament.

Range of passing from defence

108 passes

Seen higher up the pitch, esp. from
Leah Williamson (35% Of passes)

She plays in central defence (left - side)



player pitch position

Observation

A defender that pushes upfield will allow a holding midfielder to create chances in zone 14

Cluster

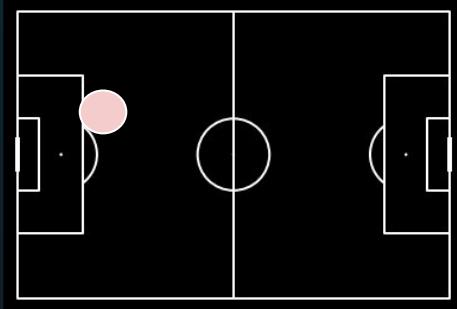
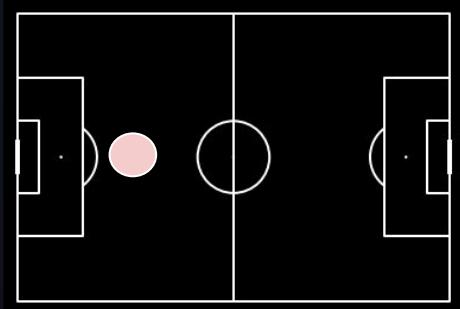
?



Keira Walsh
Holding Midfielder



Leah Williamson
Central Defender (Left)



03 Player Analysis

Pre modelling EDA

Keira Walsh's passes

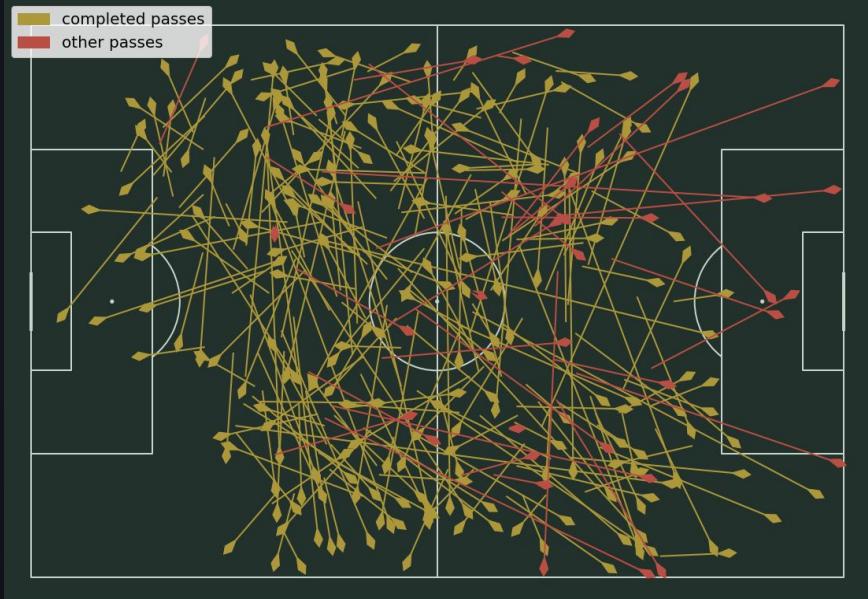
Put in no. of mins played

Passing accuracy increased

World Cup 2019

265 passes, 87% complete

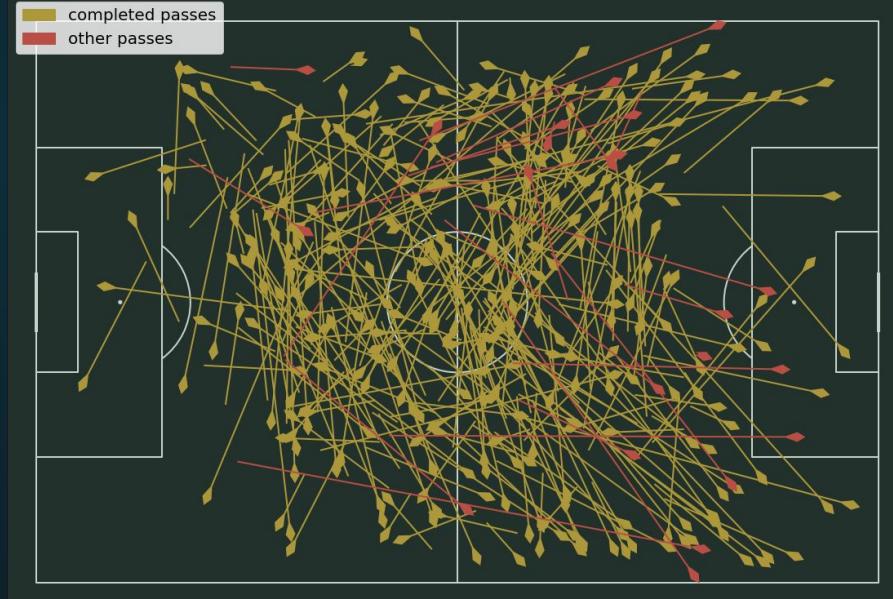
Passes



Euro 2022

339 passes, 93% complete

Passes

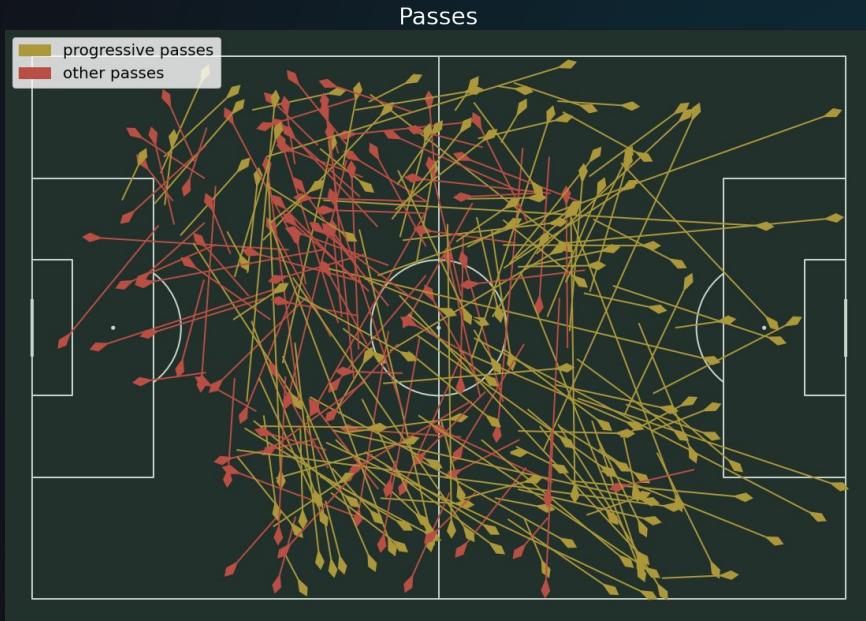


Keira Walsh's passes

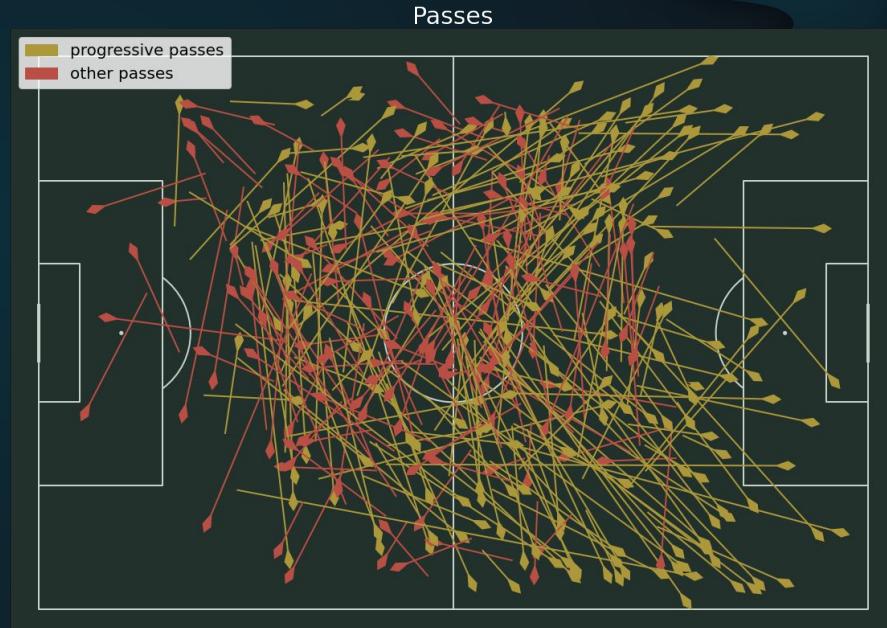
Progressive passes vs other passes

Higher progressive passing activity

World Cup 2019 - 156 progressive passes



Euro 2022 - 193 progressive passes





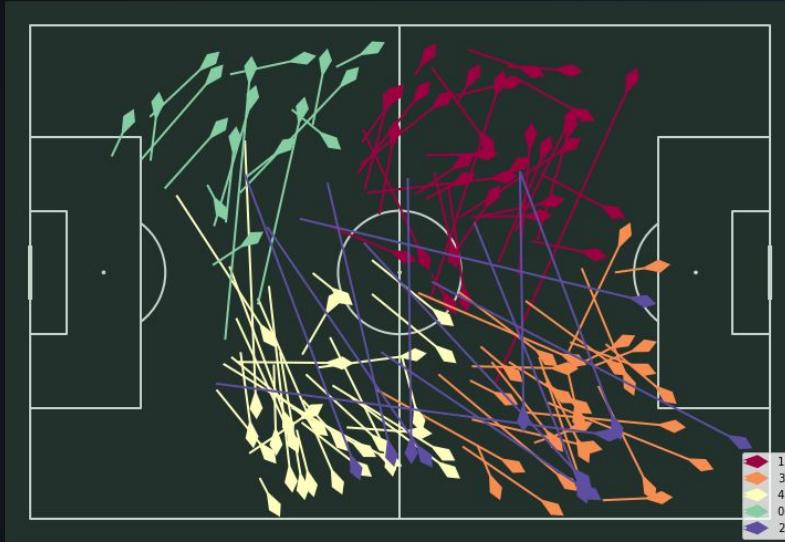
03

Player Analysis

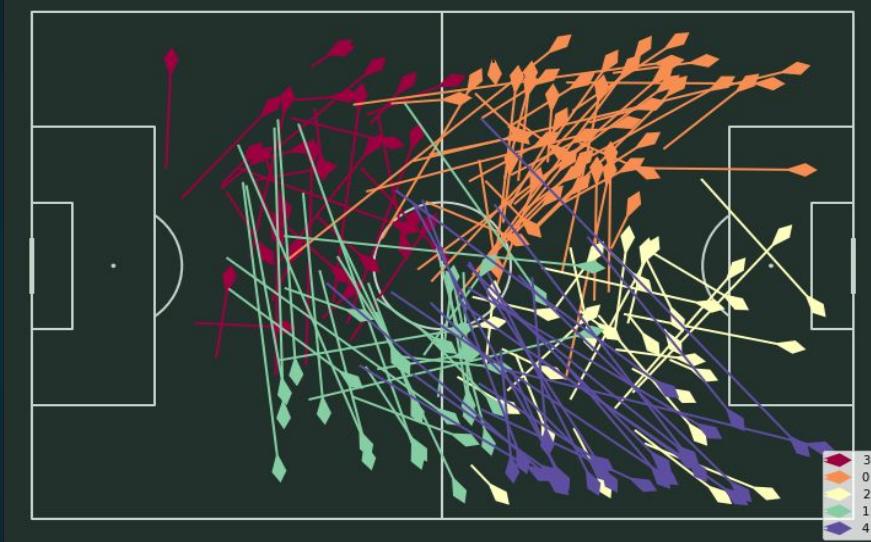
K Means & Cluster
Findings

Keira Walsh - Clustering of progressive passes

World Cup 2019 - 5 clusters
Keira Walsh



Euro 2022 - 5 clusters
Keira Walsh



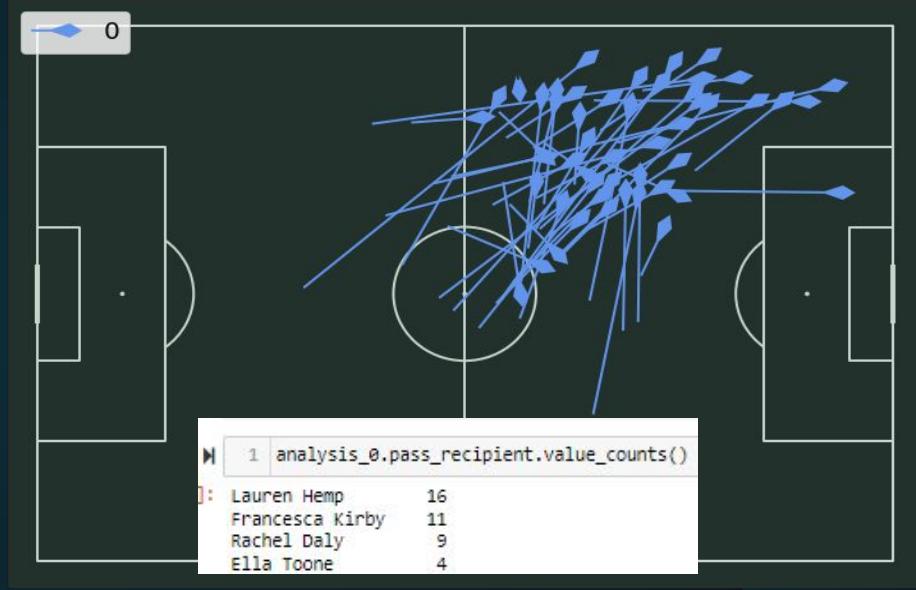
Keira's passing range have increased, with average passing length increasing in Euro 22. There is significantly more activity on the left side of the pitch in Euro 22, and with more passes reaching the opposition's 18 yard box

Keira Walsh - Clustering of progressive passes

World Cup 2019 - 34 passes,
Avg pass length 17.6 yards



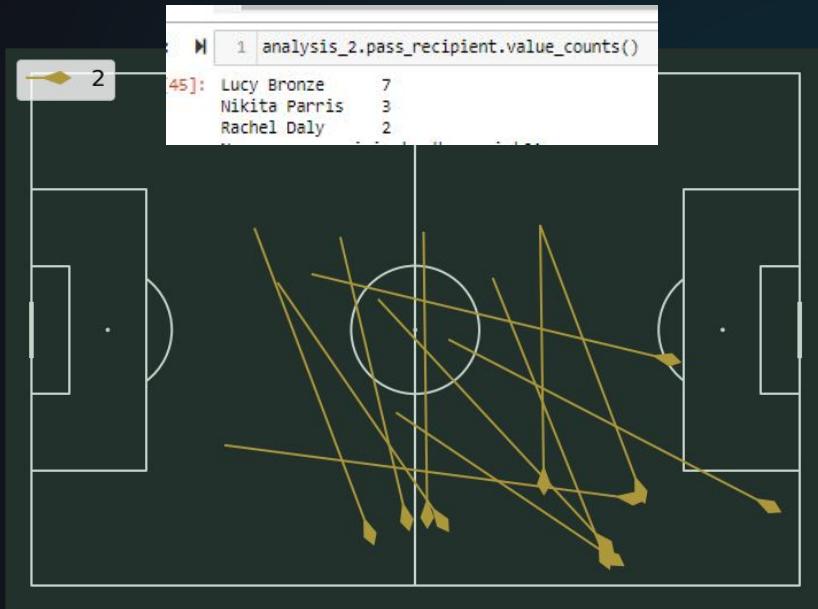
Euro 2022 - 50 passes,
Avg pass length 18.3 yards



Pass lengths were shorter on the left side back in 19, compared to longer ones in 22. This is very likely due to the speed of left wingers. In Euro 22, Lauren Hemp was running onto the left side pases, and she is one of the fastest wingers in this tournament.

Keira Walsh - Clustering of progressive passes

World Cup 2019 - 12 passes



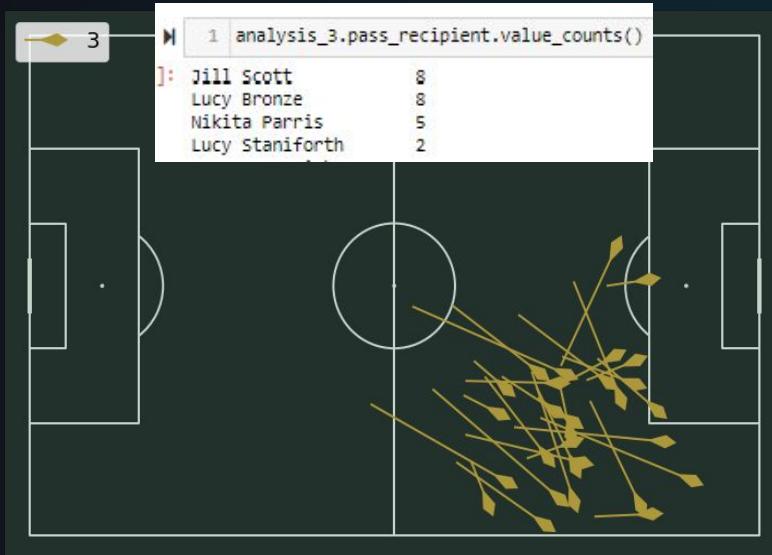
Euro 2022 - 26 passes



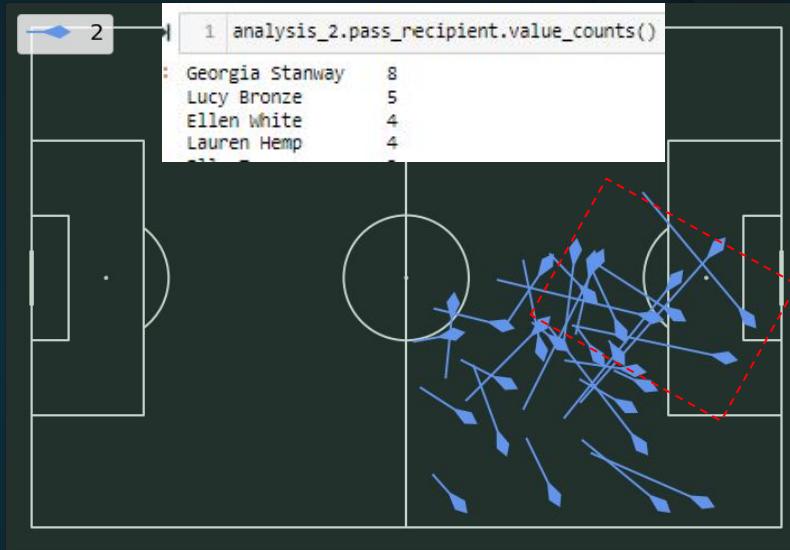
Consistency of long and high right sided passes seem to have improved, could be a mixture of better passing technique and presence of speedy right wingers in euro 22. In world cup 19, these passes were mainly made to right full back rather than a winger.

Keira Walsh - Clustering of progressive passes

World Cup 2019 - 29 passes



Euro 2022 - 31 passes



Keira seems to be able to better progress her passes into the final third in Euro 22, as compared to 19. Let's investigate further.

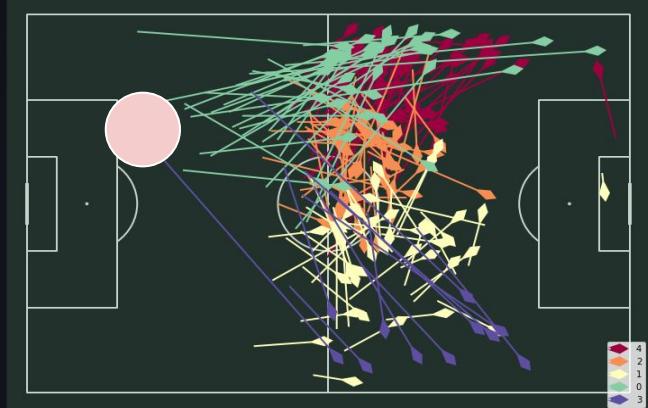
"Secret Midfield Player"

Leah Williamson

Defender, England

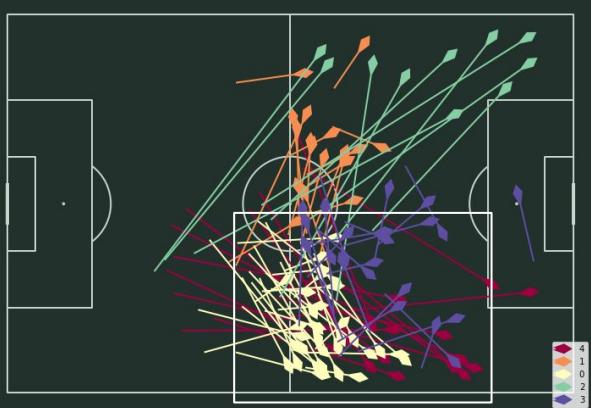
Defender, Germany

Leah Williamson



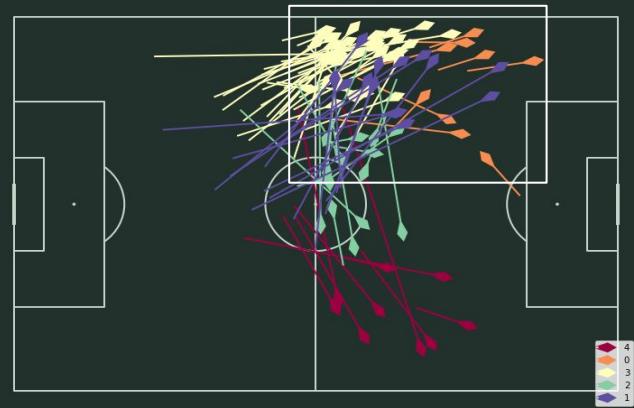
462 passes made, 41%
in offensive half

Millie Bright



298 passes made, 34%
in offensive half

Marina Hegering



267 passes made, 32%
in offensive half

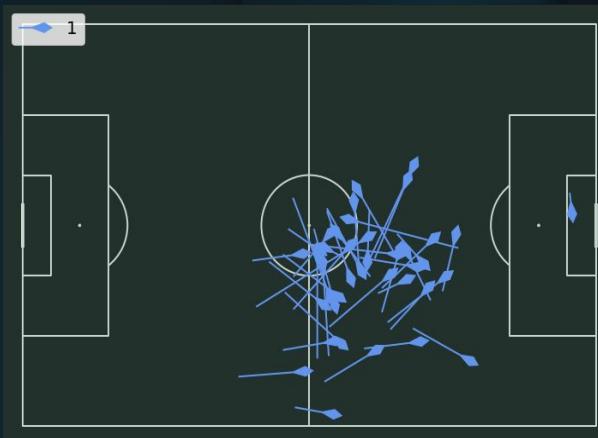
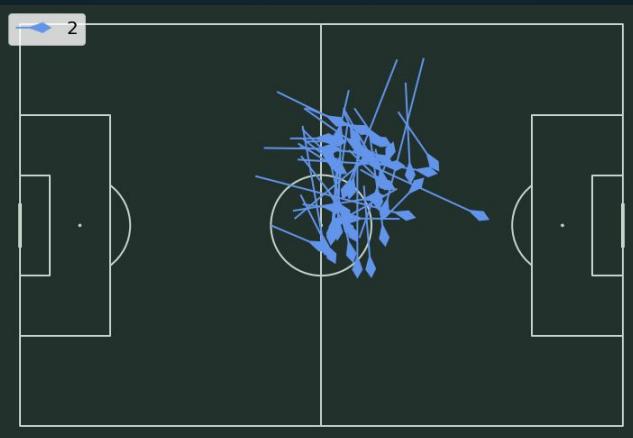
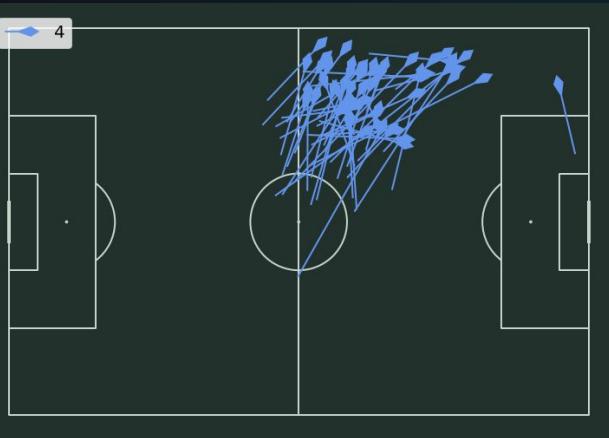
Comparing to 2 other defenders who have played the same number of games as Leah, she has made 1.5x more passes than the other 2, with 10% more passes landing in the offensive half. Her passing range is equally good from left to right, compared to other defenders who are strong in only 1 side.

Leah Williamson – Clustering of progressive passes

55 passes

50 passes

40 passes



44% passes to Rachel Daly (left back)

15% passes to Lauren Hemp (left winger)

28% of these passes go to Keira to allow her to utilise her passing range to get the ball to the wingers / forwards. The rest mostly go to attacking midfield, enabling England to launch attacks quicker.

04 Considerations & Recommendations

Some key points on England's Euro 22 success

Keira Walsh

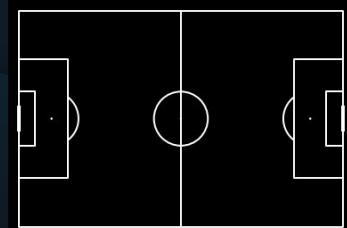
A holding midfielder with a good passing technique is critical to the success of the team, as she is able to have attacking options from the back.

Keira's technique have complemented well with fast wingers who can run onto passes. During transition play, this counter attack will generate lots of chances in the women's game where defensive pressing styles is not prevalent in most teams, unlike the men's game.

Leah Williamson

A defender who can push up forward and pass with precision, instead of just clearing balls and doing short and backward passes, will have a player advantage in midfield, allowing more players in attacking positions to create chances.

On the flip side, this style of playing requires a high defensive line as Leah plays up to center line sometimes. This is susceptible to counter attacks if the other defender isn't speedy or the defensive positioning isn't good enough.



Considerations

- Opposition tactics need to be taken into account as explanation on why England play the way they do.
- Data like through passes on goal, passing sequences leading to goal chances, can be incorporated into the clusters to understand where England's goal threats and opportunities come from

THANKS!

Does anyone have any
questions?