UNIQUENESS PASSING MODEL

by FC_rSTATS

FC rSTATS

01. INTRODUCTION

You are watching a football match, possession is being circulated around the middle in unspectacular fashion. Then all of a sudden, your central midfielder (known for his vision) gets a moment of time and hits a great pass. You know it's a great pass, it's something different, yeah it's not normal, it's..... unique

I have attempted to capture this ‘uniqueness’ in a metric, in the hope that it can become a reliable inclusion in player recruitment shortlisting and analysis

02. THE DATA

I will be using the recently released World Cup 2018 dataset from Statsbomb. The dataset consists of over 63 games which totals to 174,418 events, of which 56,943 are passes.

Further Reading

Statsbomb Resource Centre
Sign-up to their public data user agreement, get access to the data & get more info on the dataset

StatsBombR Package
The R package that will help you get your hands on that data super fast and super easy

The Statsbomb Blog
The epicentre of public analytics, great writing and includes examples of how the free data could be utilised

Accounts to Follow

Ted Knutson
Founder / CEO
@mixedknuts

Mike Goodman
Managing Editor
@TheM_L_G

Thom Lawrence
Tummyball Chief
@deepxg

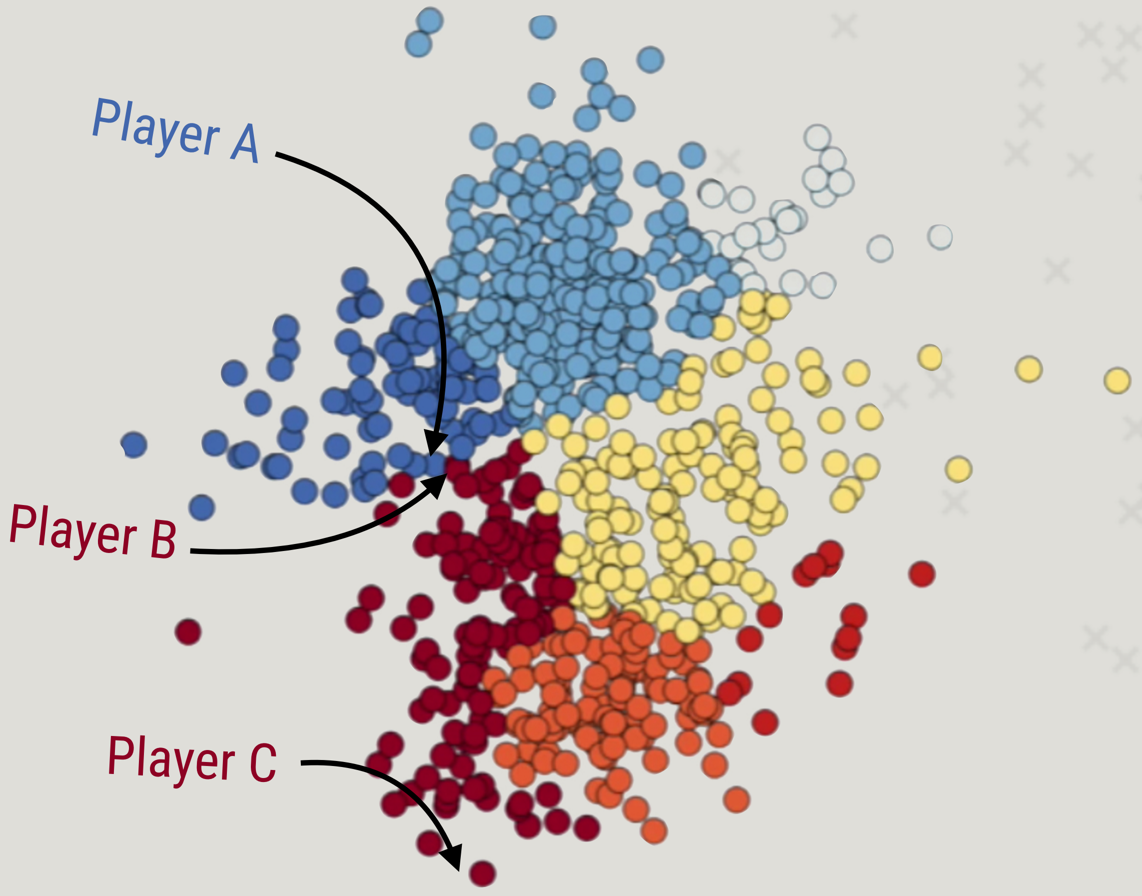
James Yorke
Analyst and More
@jair1970

03. CLUSTER ANALYSIS

Clustering allows observations (rows) of data to be compared across a dataset and to be grouped or *clustered* together by their similarity.

The technique has been widely used due to ease of use, it's ability to improve comparative analysis and it's intuitive nature. It has it's strengths, as well as it's weaknesses.

For example, in the representation of multi-dimensional space, Player A and Player B are close to each other but it in different clusters. Whilst Player B and Player C are more dissimilar but are clustered together.



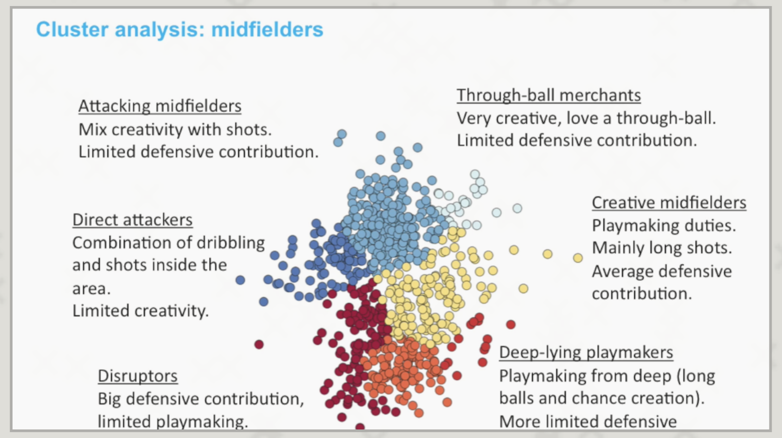
Further Reading

Introduction to Various Clustering Techniques

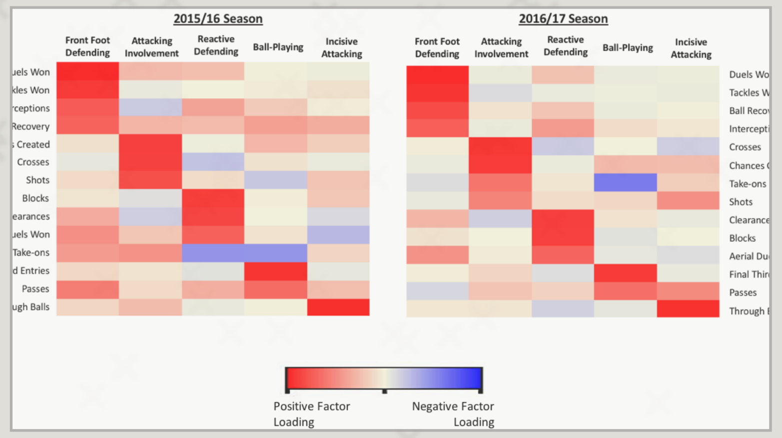
Introduction to K-means Clustering

Introduction to Principal Component Analysis

Identifying & Assessing Team Strategies
Will Gürpınar-Morgan
Use of k-means clustering to group possession chains, in order to identify team strategies



Square Pegs for Square Holes
Will Gürpınar-Morgan
Use of k-means clustering to identifying player types to support an underpinning recruitment strategy



The Evolution of the Full-Back
Mark Carey & Mladen Sormaz
Use of PCA and clustering to examine the development of the full-back and group into types

04. CALCULATING DIFFERENCE

Leaning on the work of clustering, I will use the Nearest Neighbour Search (NNS) function to calculate the 'uniqueness' of each pass

The NNS will calculate the dissimilarity of pass *n* with other passes as the Euclidean *distance* within the multi-dimensional space of the data

Instead of creating clusters, the 'uniqueness' metric will be the sum of the *distance* of a certain number of passes. To make this analysis scalable for more data, 1% of the total passes is used as the comparison scope (569 passes in this case)

The higher the *distance* the more 'unique' the pass is deemed

Each pass will be compared for it's uniqueness across the following characteristics:

- **duration:** the time the pass took from release to receive
- **pass.length:** distance from origin to destination
- **pass.angle:** the angle of the pass measured in radians
- **pass.height.id:** 3 grade variable for pass height (1) ground pass (2) mid-height (3) high-ball
- **x:** the vertical location of the pass origin
- **y:** the horizontal location of the pass origin
- **x.destination:** the vertical location of the pass destination
- **y.destination:** the horizontal location of the pass destination

05. DOES IT MATCH?

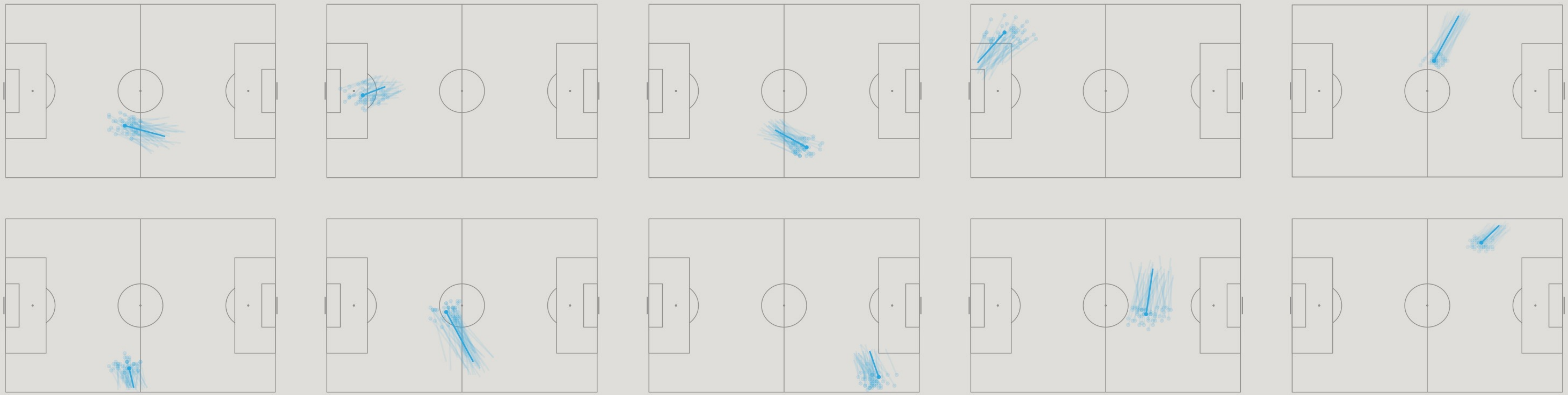
I grab a random selection of 10 passes to check the nearest neighbour search is doing it's job and finding the 50 most similar passes. Does it match? Yes, it's doing a very good job at matching similar passes, so let's delve a little deeper into the results

Visualisation Explainer

Pass Origin - Circle denoting the origin of the pass

Pass Path - Line denoting the path of the pass

Similar Passes - Lower Opacity for matched passes



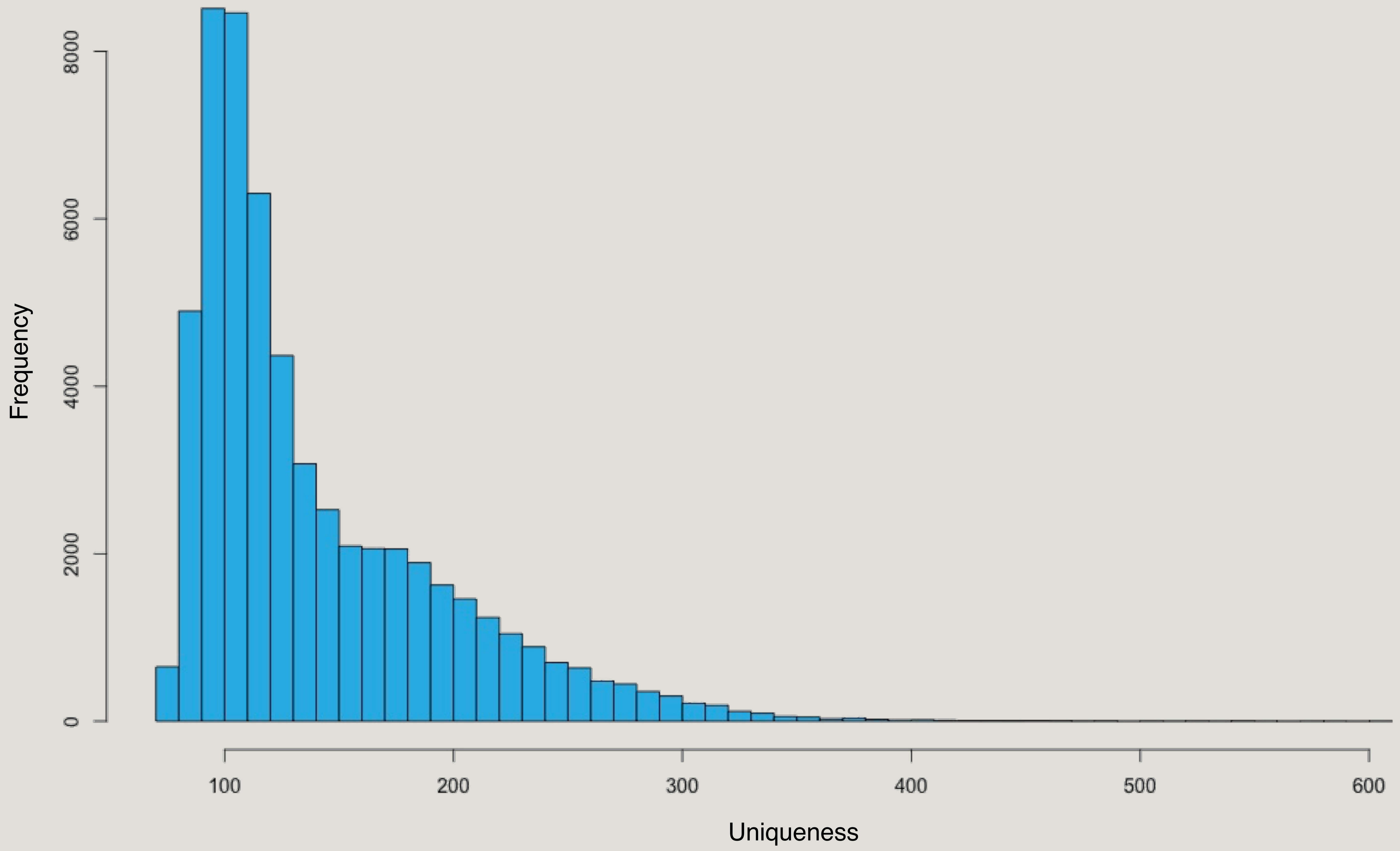
06. HISTOGRAM OF UNIQUENESS

The **histogram** shows a Right-Skewed / Positive Skewness distribution of uniqueness

Average Uniqueness 140.2

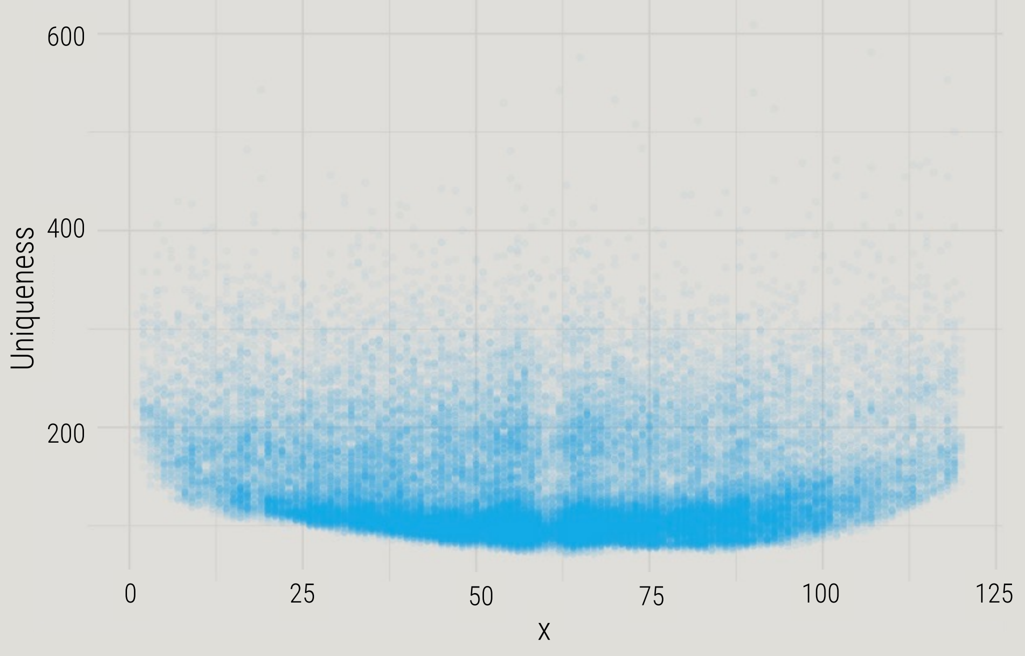
Quantiles
0 % 70.4
25 % 100.2
50 % 119.4
75 % 168.8
100 % 608.9

The positive skew extends to the extreme of 608.9, this could mean that there are some really unique passes that could offer a team real value



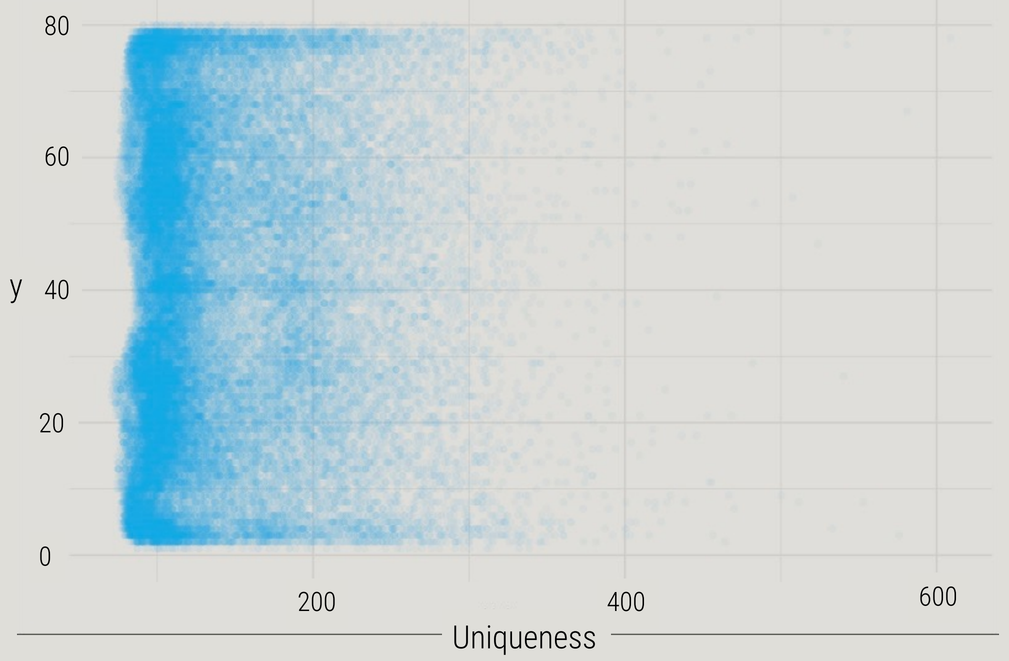
07. INITIAL DATA EXPLORATION

x Location of Pass Origin



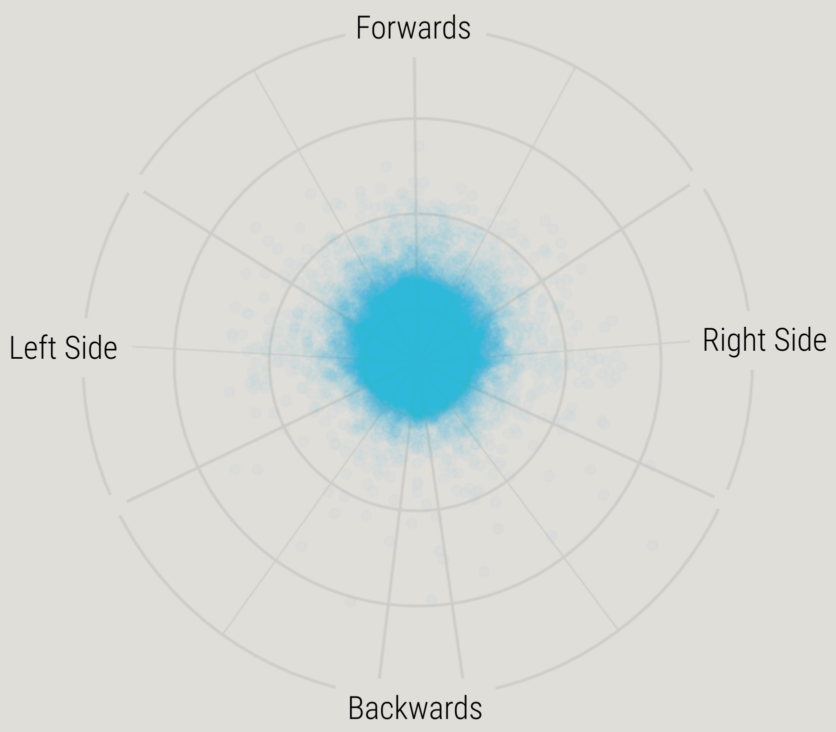
There is no specific and clear pattern other than a mild curved dip. x location of the pass origin does not seem to be overly descriptive

y Location of Pass Origin



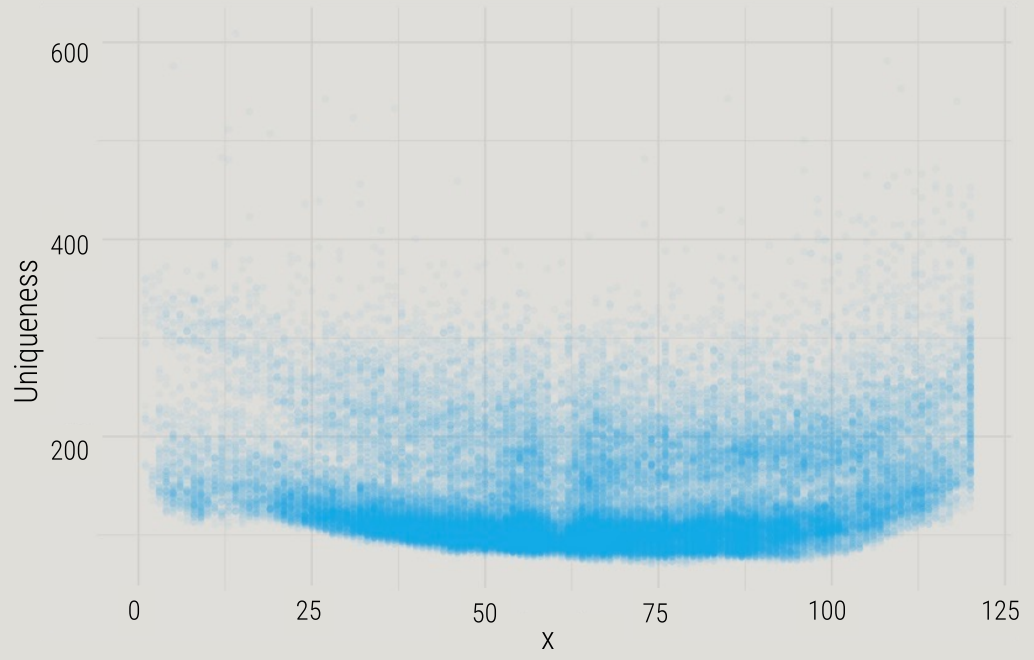
Once again is no specific and clear pattern other than a Uniqueness reaching high values in at the wings

Pass Angle



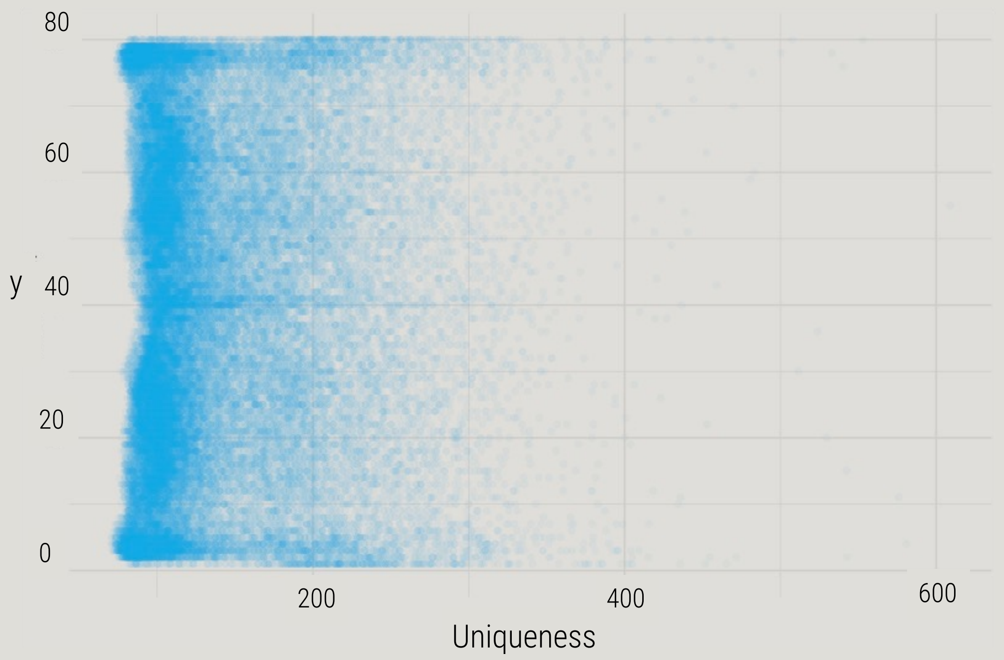
There is no specific pattern between pass angle and uniqueness aside from a high values being seen in forward passes

x Location of Pass Destination



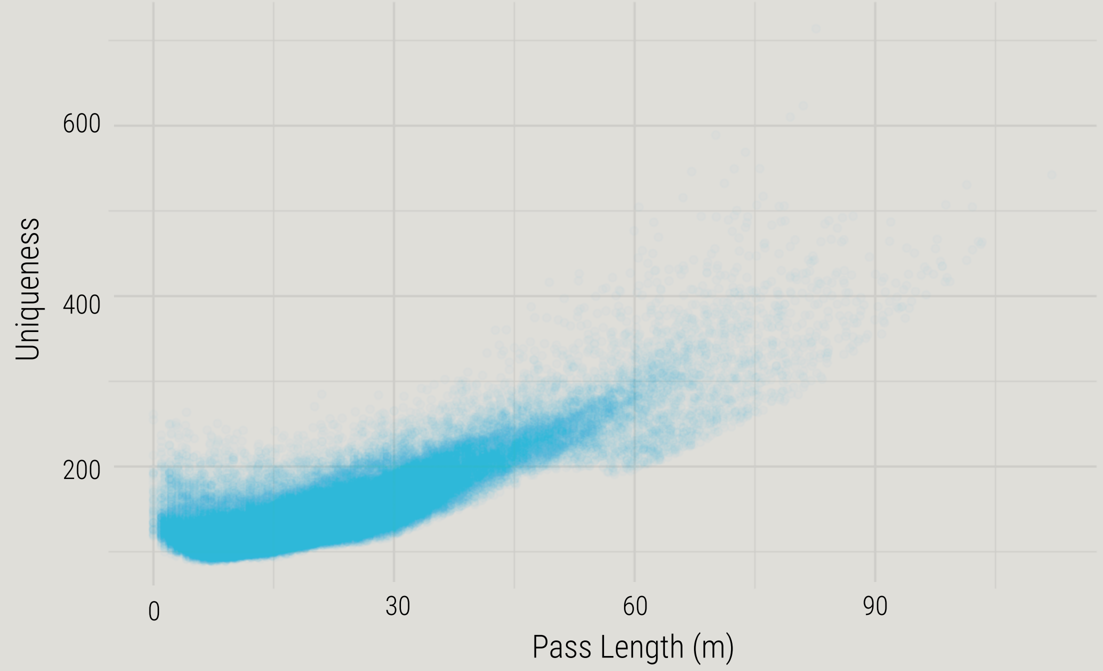
Once again is no specific and clear pattern other than a mild curved dip and a distinct uplift in very advanced receives.

y Location of Pass Destination



Once again is no specific and clear pattern other than a Uniqueness reaching high values in at the wings

Pass Length

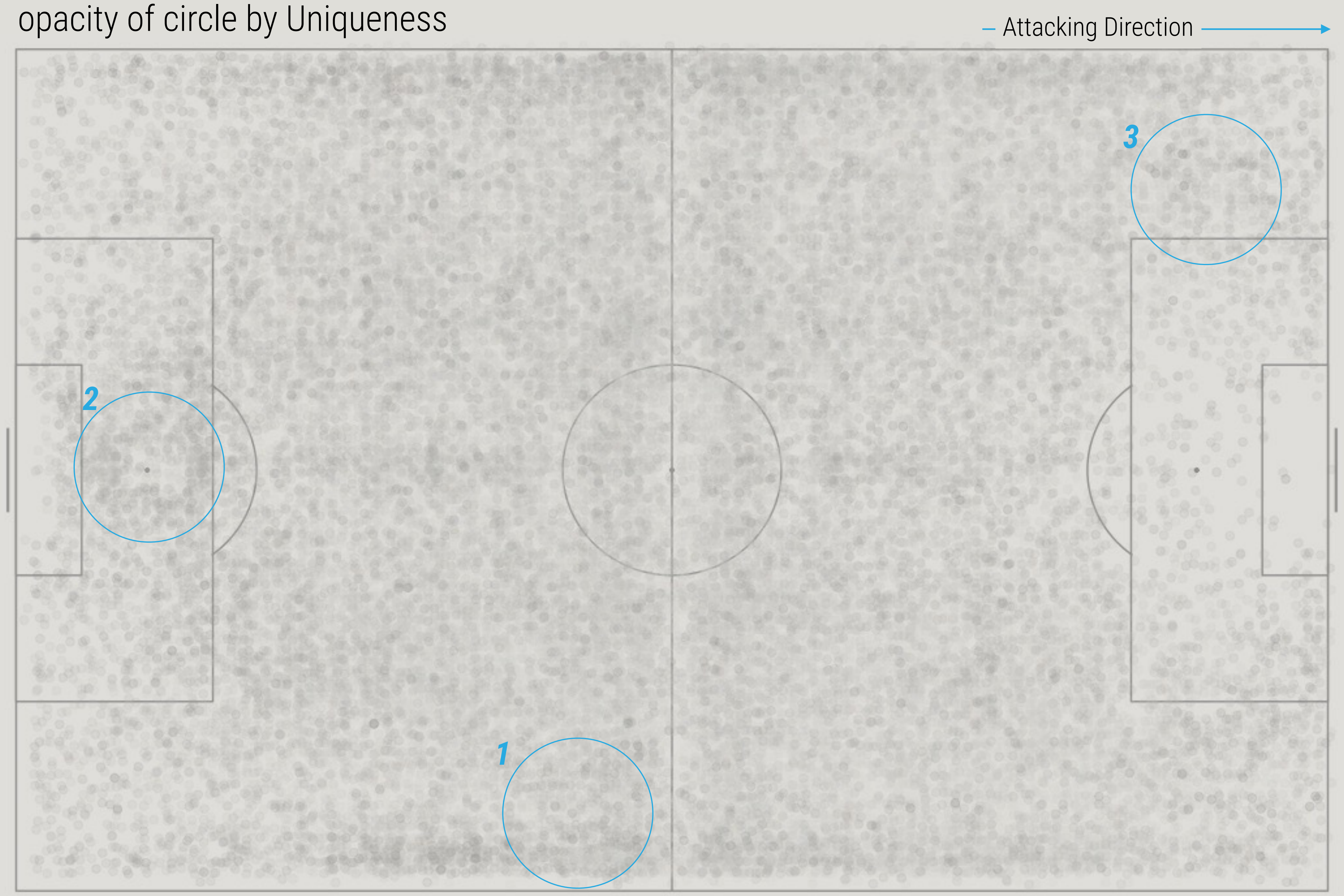


There is a distinct pattern in pass length with the longer the pass the higher the uniqueness rating

08. LOOKING FOR PATTERNS

x,y Location of Pass Origin

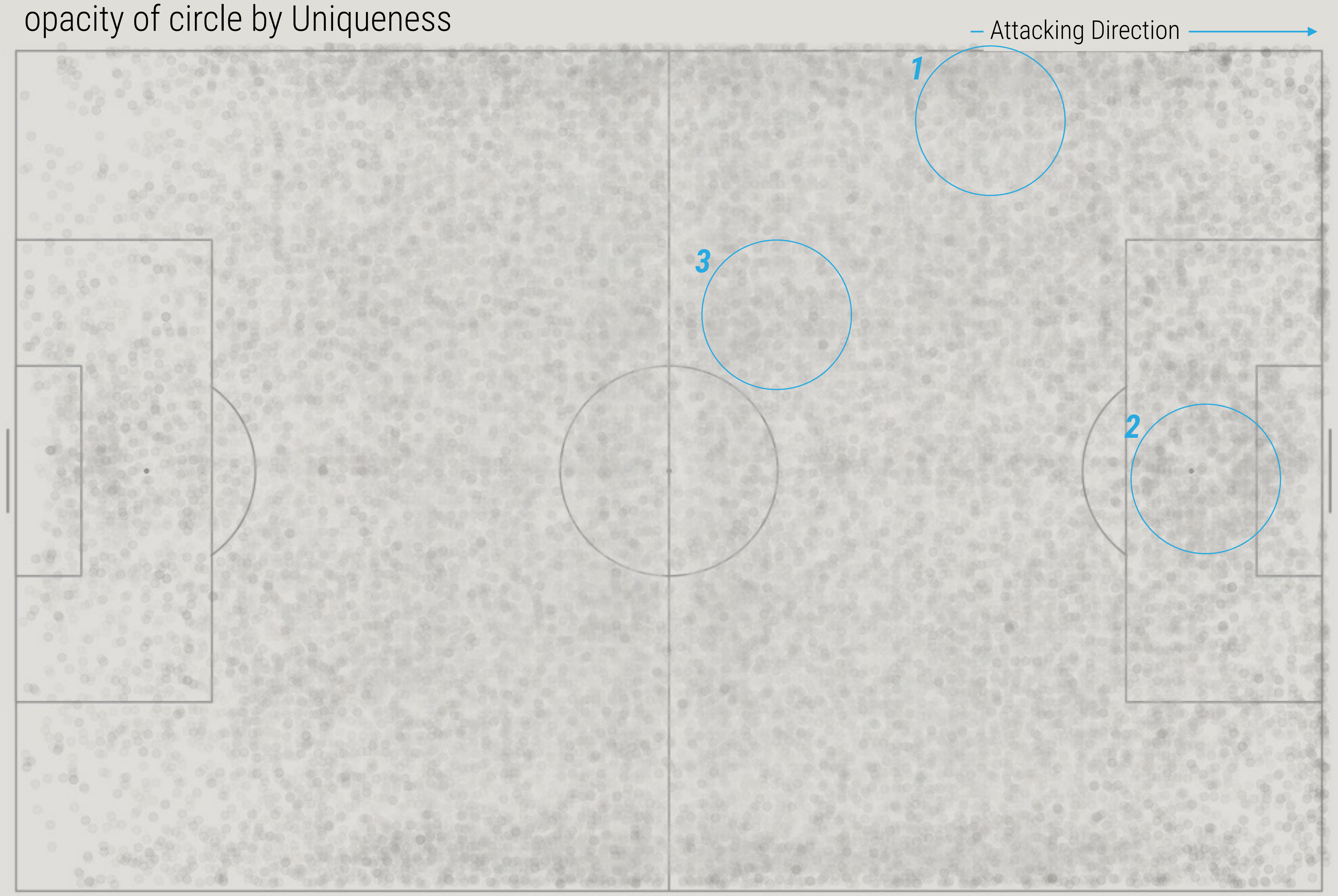
opacity of circle by Uniqueness



1. An increase at the edges of the pitch
2. An increase in the defensive box
3. An increase in wide areas around the box

x,y Location of Pass Destination

opacity of circle by Uniqueness

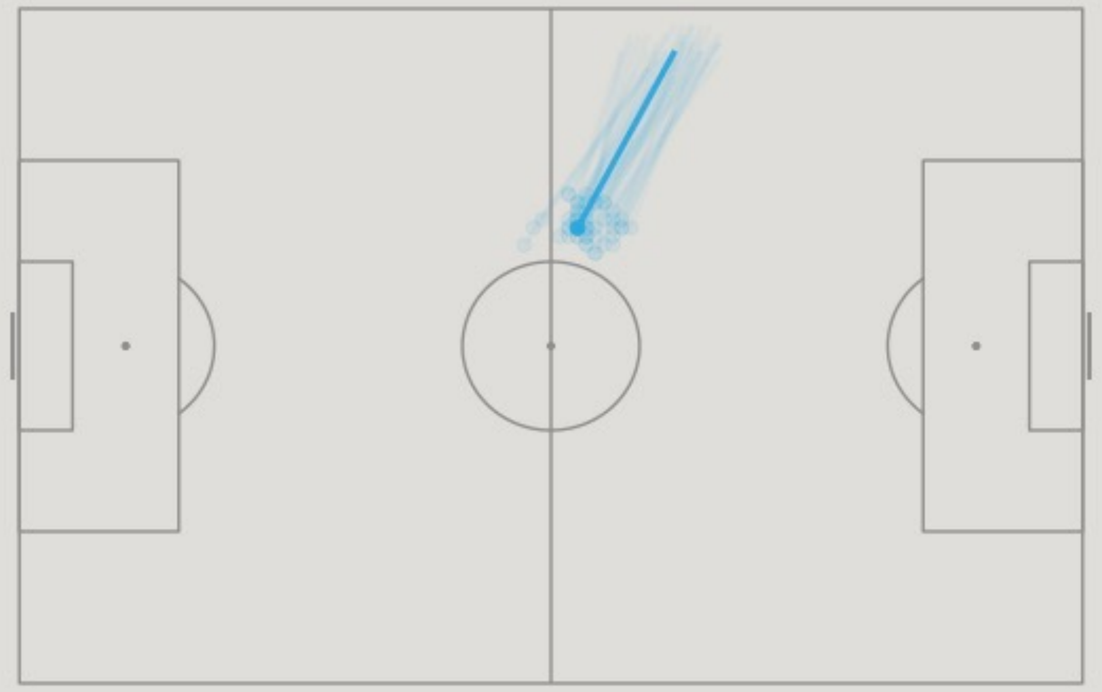


1. An increase at the edges of the pitch
2. An increase in the offensive box
3. An interest deep lying play-maker spot?

09. WATCH THE VIDEO!

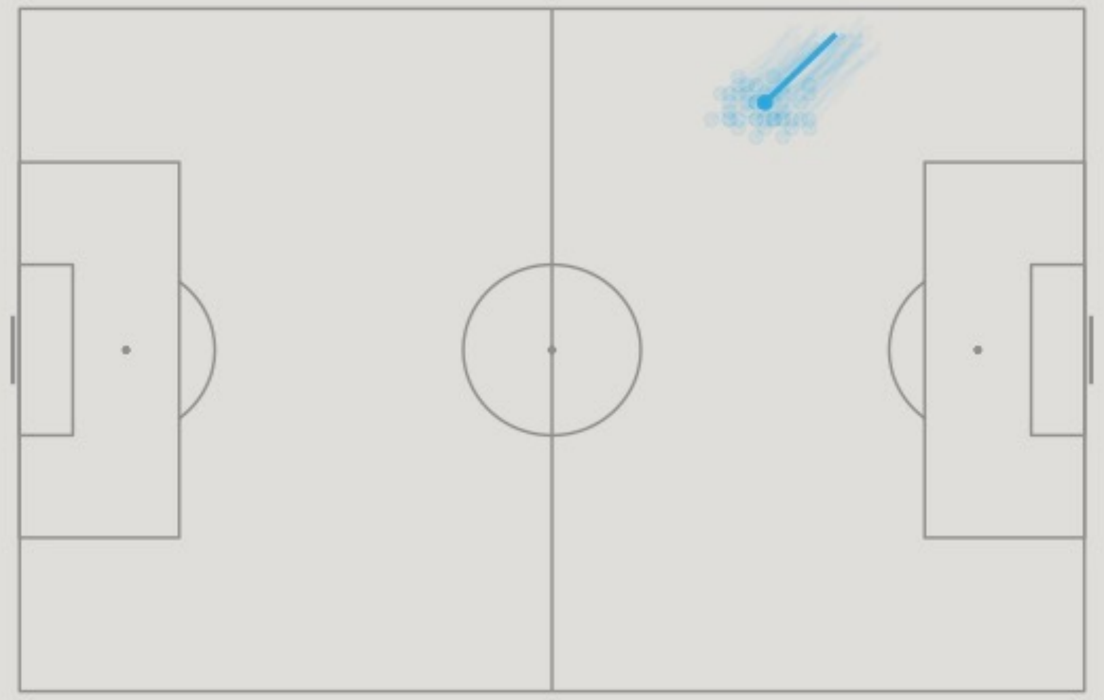
Sergio Ramos
v Russia
A simple slow pass out to Jordi Alba, this is the least unique pass in the whole dataset

Uniqueness = 70.5



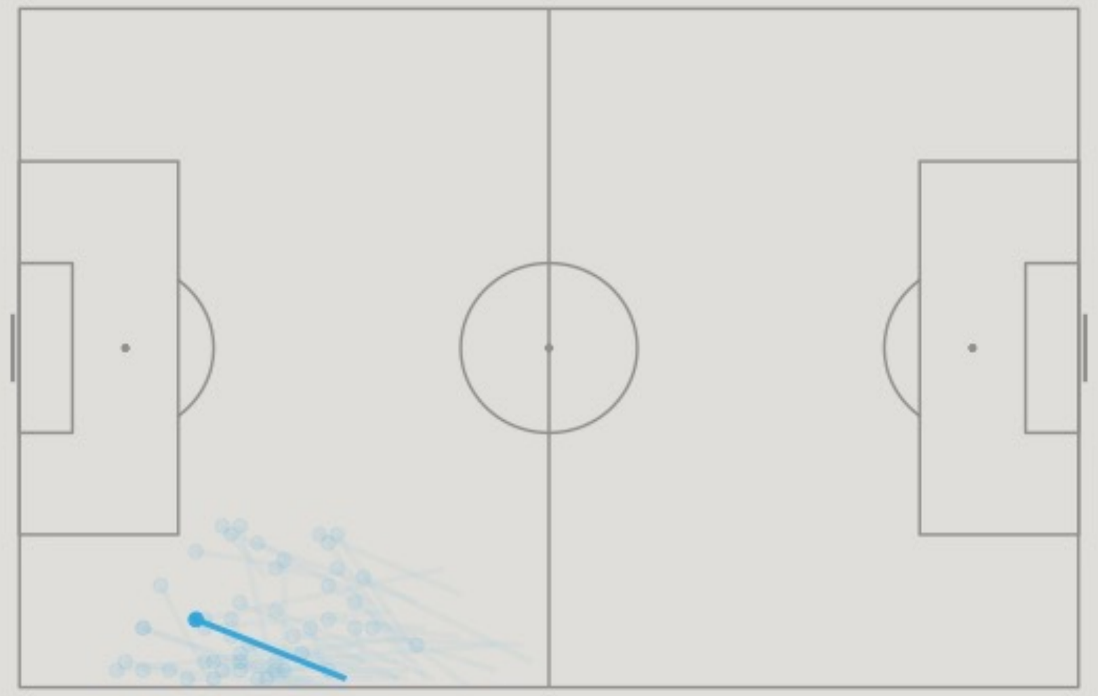
Isco
v Russia
A simple slow pass and short pass out to the wing. A continuation pass, which is surprisingly common

Uniqueness = 74.0



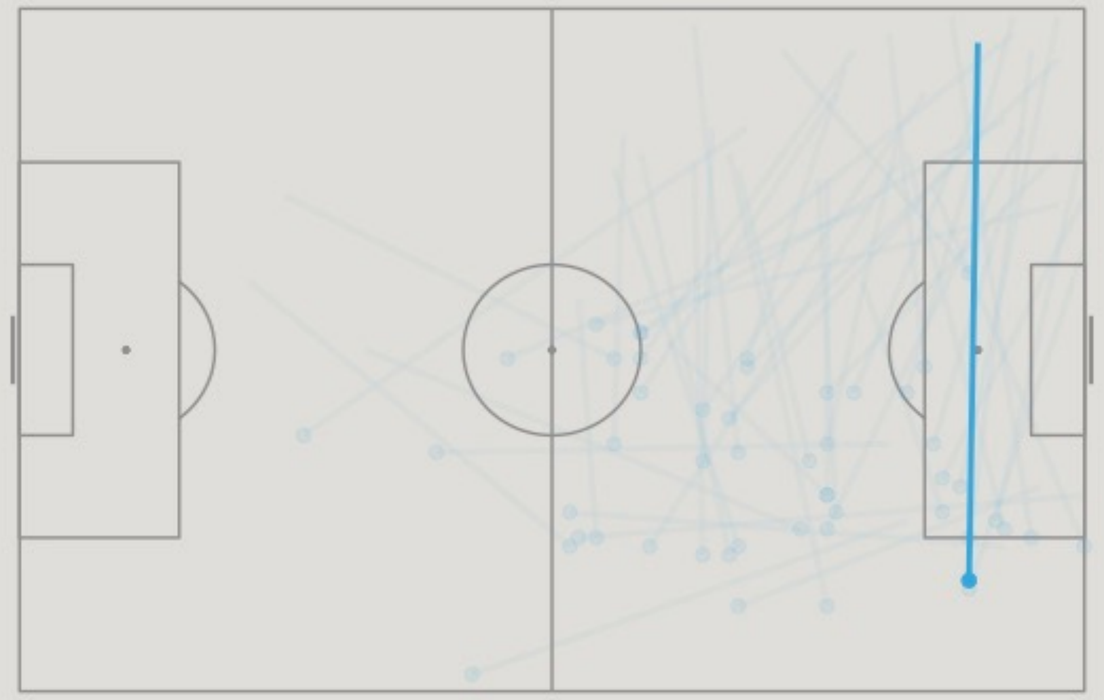
Simon Kjær
v Croatia
A surprisingly unique pass, these situations are rarer than I imagined and have a low conversion rate

Uniqueness = 330.4



Paulinho
v Costa Rica
A great pass in theory but as it overrun it became highly unique, in fact the most unique pass in the dataset

Uniqueness = 580.9



10. DON'T WATCH THE VIDEO!

In my hypothesis, the more unique a pass is the better is it. Paulinho's pass vrs. Costa Rica made me revaluate the passes with the highest uniqueness rating.

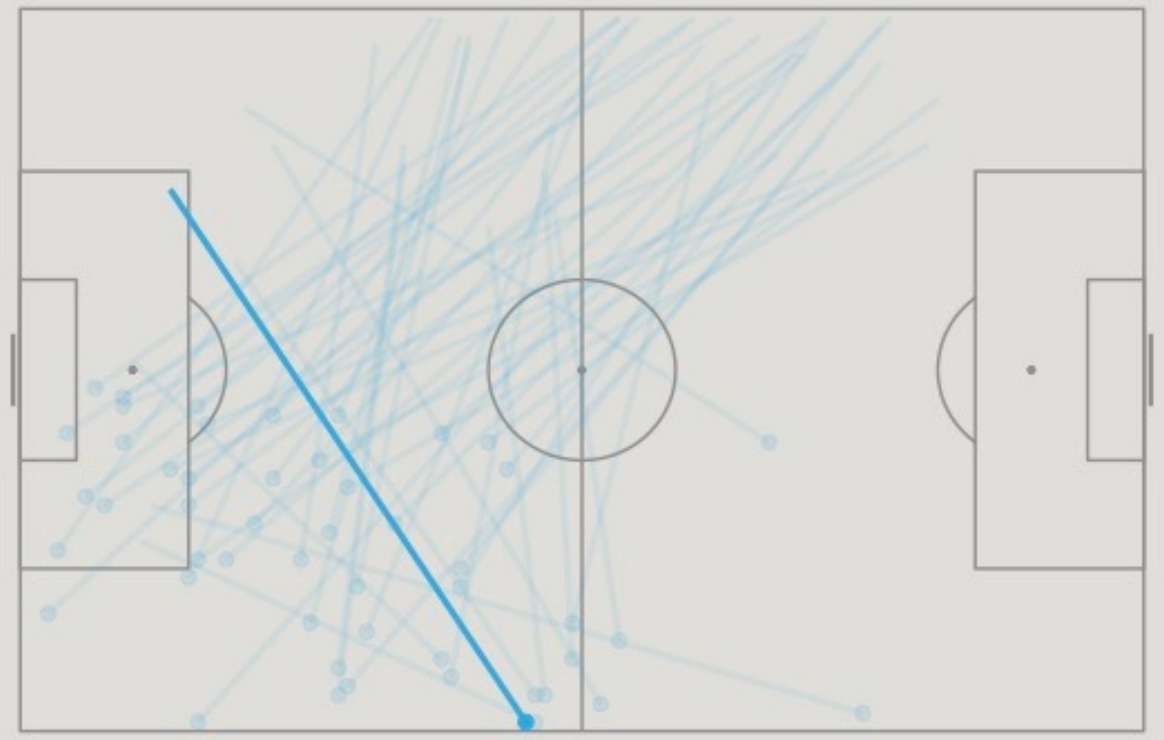
I am starting to have a feeling that the higher values are in fact so random that they score highly but are in fact not of much use and a bad decision to attribute to a player as a positive.

Let's have a look at some other passes with high uniqueness rating

Carvajal
v Morocco

A cross field ball, awkward to receive by Pique which resulted in the ball running on further. Certainly not an interesting pass!

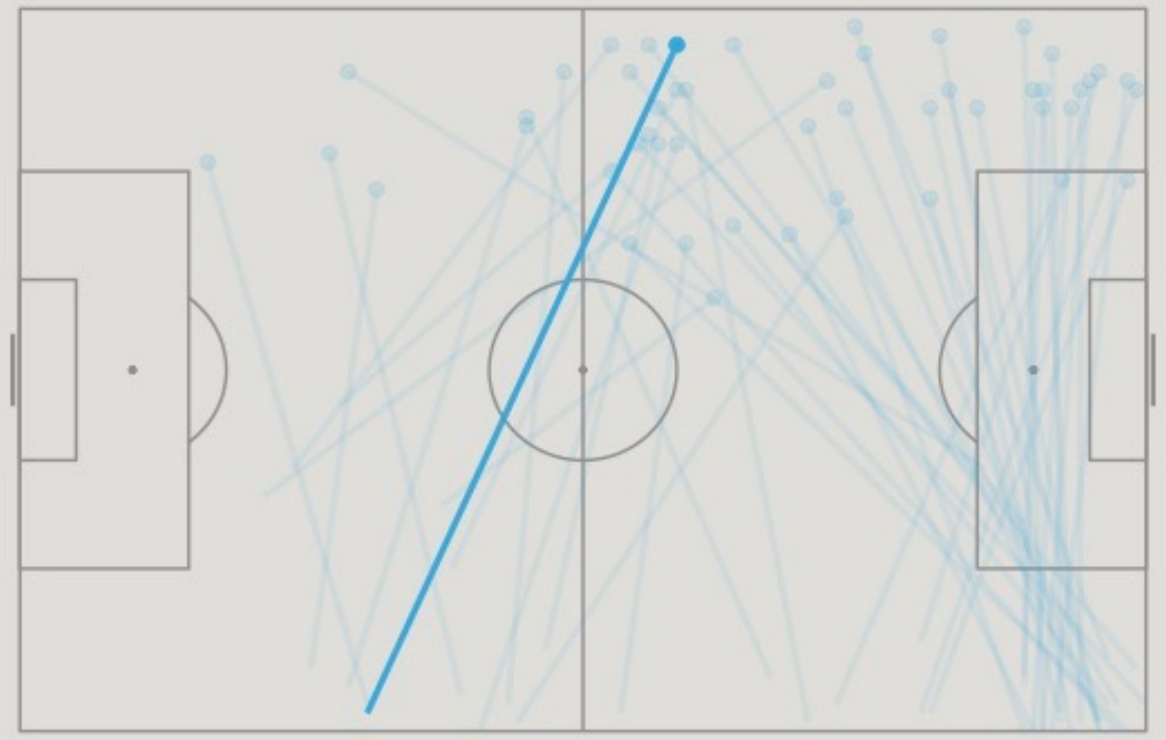
Uniqueness = 529.6



Di Maria
v Nigeria

A very strange cross field volley under slight pressure. Possible intent to play to central defenders but overhit. Again, not an interesting pass!

Uniqueness = 533.00



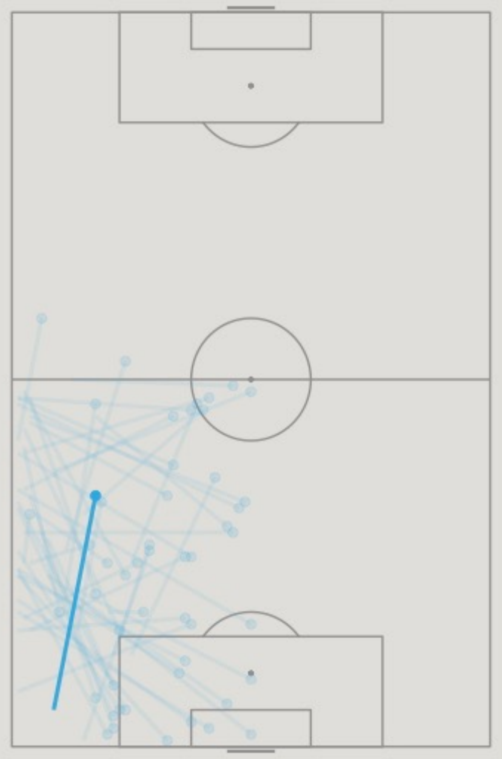
11. A METRIC CRITIQUE

1. The passes with the highest scores of uniqueness are very random. By leaving these passes in for further analysis players and teams will be rewarded.

Having studied a vast number of these passes it seems that the top 0.25% of passes are totally random and can be screened out.

The top 0.25 - 1.00% does still include some random passes but also includes those with real quality. These will be left in the dataset, yet I must retain the understanding that this is a weakness of the model.

2. There's high prevalence of passes in the defensive half that rank highly on the uniqueness scale. A pass to a full-back who dropped deep & wide is rare and high in uniqueness



Although these passes are indeed unique they didn't inspire the thinking behind the metric! When I run further analysis on the results defensive players dominate all of the shortlists. I will therefore filter the passes, to create a subset of *unique attacking passes.*

x > 60 & x.destination > x *or* x > 80 & x.destination > 60

3. All passes have a unique score. If we sum the scores in our further analysis player making 10 very boring and simple passes would score higher than a player making 1 truly amazing pass.

Having studied the video of a vast number of these passes it seems that the top 25% of the unique attacking passes seem to be those that the metric was intended to identify.

Therefore I keep just these passes in the classification of *unique attacking passes* – leaving us with 5,410 passes

12. SHORTLISTING 'UNIQUENESS'

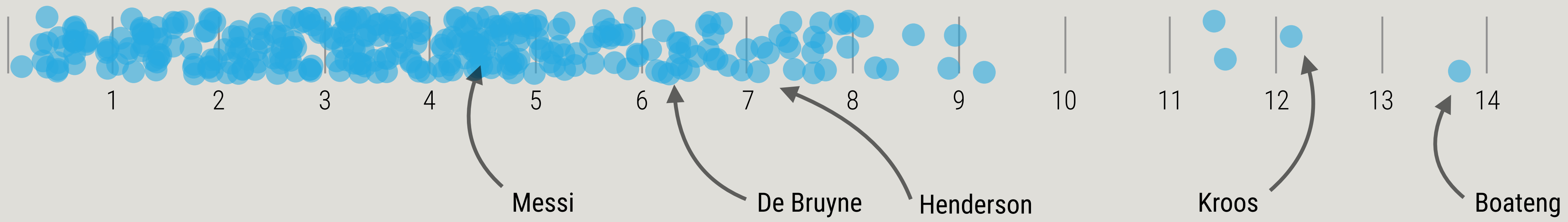
Shortlisting provides us with some insights via an eye-ball test of who pops up on the lists.

These shortlists would become much more accurate after 20+ games. I have applied a filter of >10 UPA & >180 total minutes played.

The shortlists largely pass the eye-test. Toni Kroos tops the pile, high possession teams like Germany and Spain show up and and Red Bull Leipzig features twice.

Yasir Al Shahrani and Tarek Hamed are names that I don't have good knowledge of. Yet when reviewing their footage I can understand why they are scoring highly. For example ——>

UPA PER 90: Overall Distribution

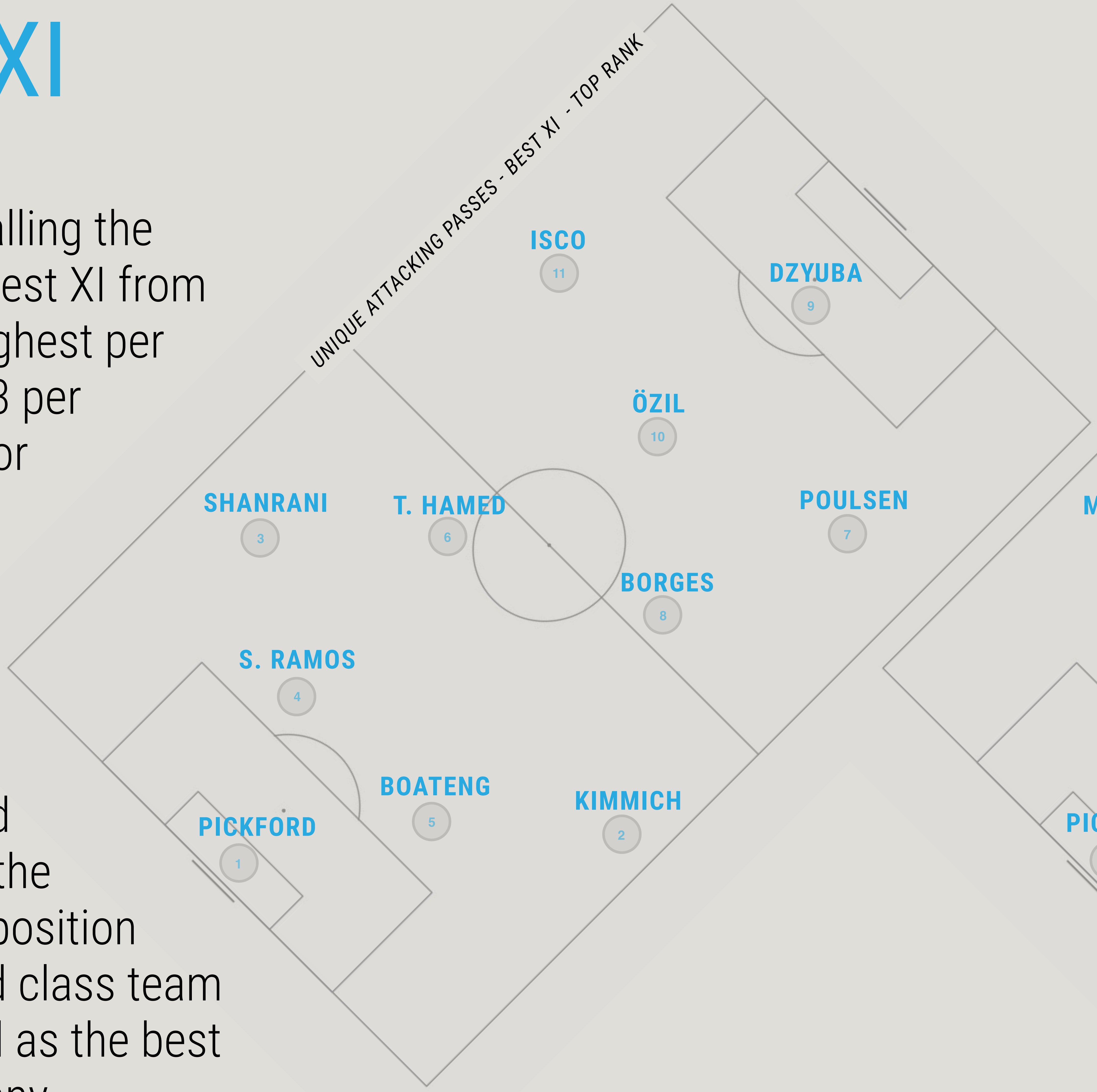


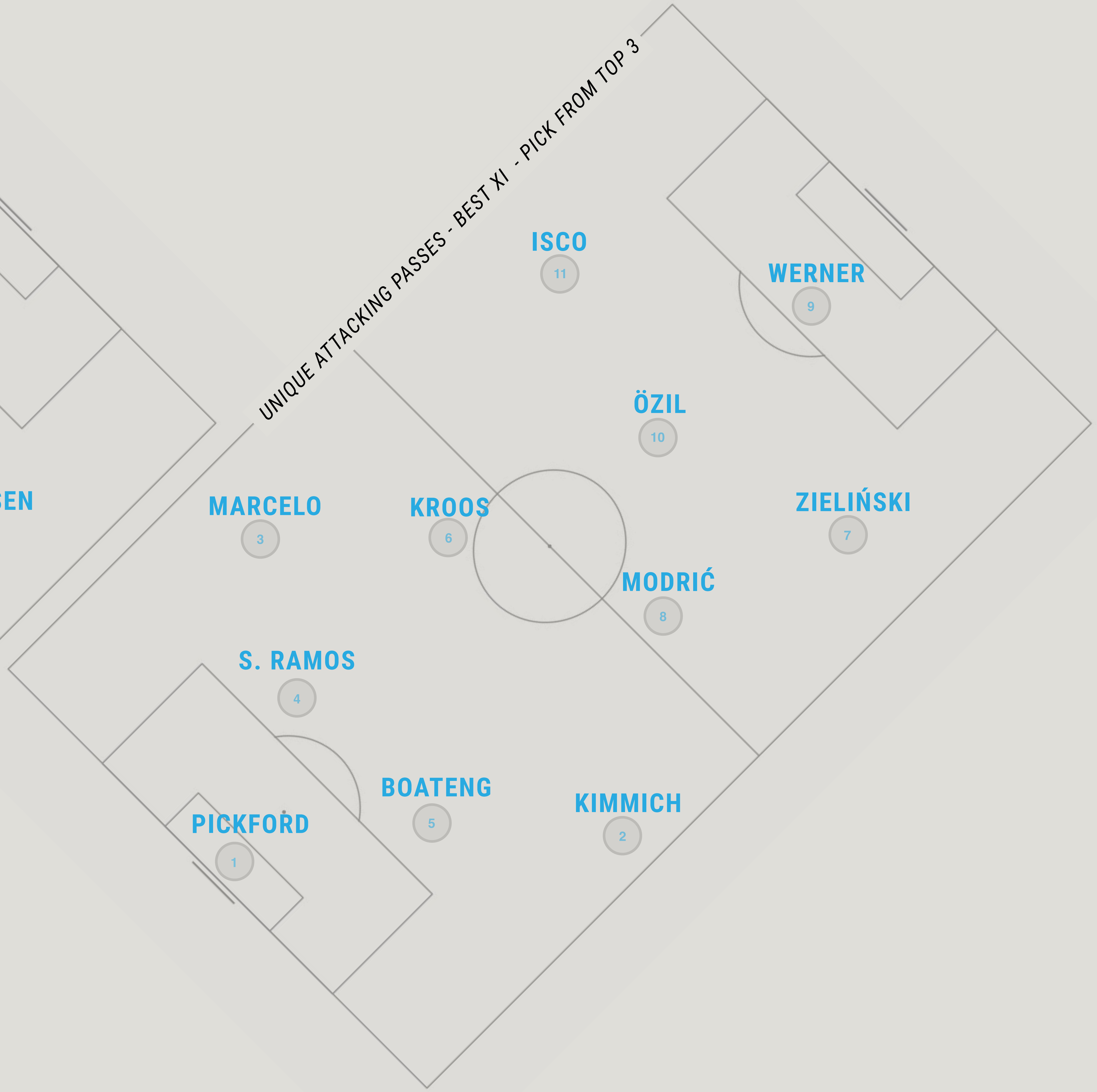
PLAYER	TEAM	CLUB	UPA PER 90
			Defenders
Jérôme Boateng	Germany	Bayern Munich	13.7
Joshua Kimmich	Germany	Bayern Munich	11.5
Sergio Ramos	Spain	Real Madrid	7.7
Yasir Al Shahrani	Saudi Arabia	Al-Hilal Riyadh	7.6
Marcelo	Brazil	Real Madrid	6.6
			Midfielders
Toni Kroos	Germany	Real Madrid	12.1
Isco	Spain	Real Madrid	9.2
Mesut Özil	Germany	Arsenal FC	8.9
Tarek Hamed	Egypt	Zamalek SC	8.6
João Mário	Portugal	Inter Milan	8.2
			Forwards
Artem Dzyuba	Russia	Zenit St. Petersburg	11.4
Timo Werner	Germany	Red Bull Leipzig	9.0
Yussuf Poulsen	Denmark	Red Bull Leipzig	8.3
Piotr Zieliński	Polamd	Napoli	8.1
Thomas Müller	Germany	Bayern Munich	7.9

13. BEST XI

Another way of eye-balling the findings to is form a Best XI from the players scoring highest per position or in the top 3 per position (to account for tight calls)

Although the Top RankXI has some players that you would change, picking from the top 3 players in each position gives you a truly world class team full of players deemed as the best in their position by many





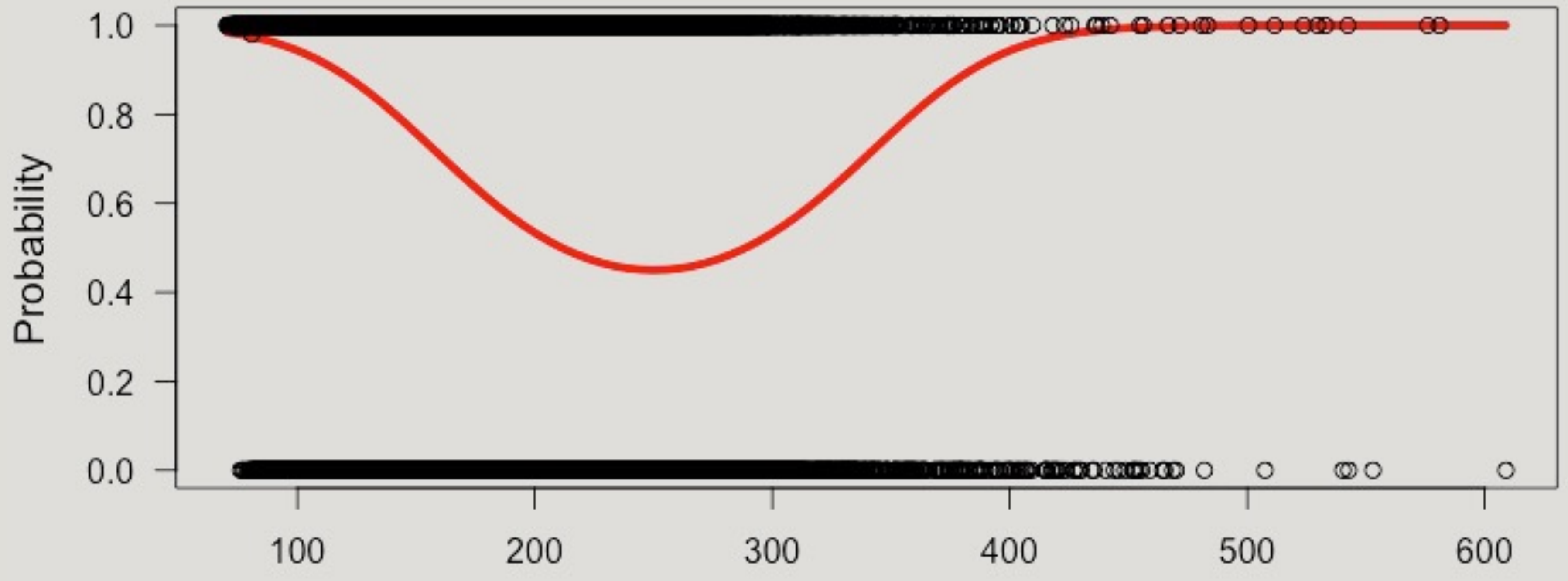
14. EXPECTED COMPLETIONS

The natural extension of identifying the uniqueness is to see if the metric or the process can help build a more precise pass completion analysis.

Pass completion stats are often see as largely irrelevant but they are intuitive so if the stat can be improved it's a big bonus.

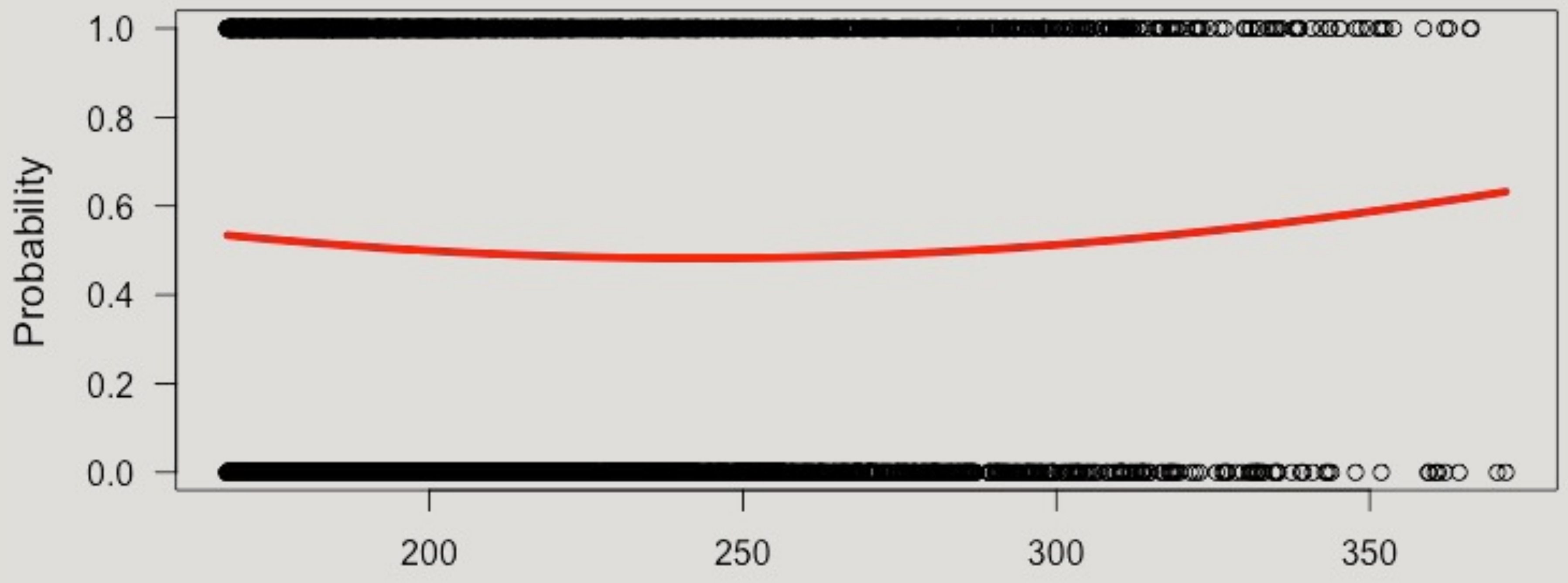
Let's quickly investigate how the uniqueness of a pass impacts on pass completion rates

Pass Completion Probability
by Uniqueness - All Passes



There is a clear dip in completion probability within the mid 200s to a low of 0.47

Pass Completion Probability
by Uniqueness - Unique Attacking Passes



A slow rise in completion probability towards the upper end of the uniqueness range

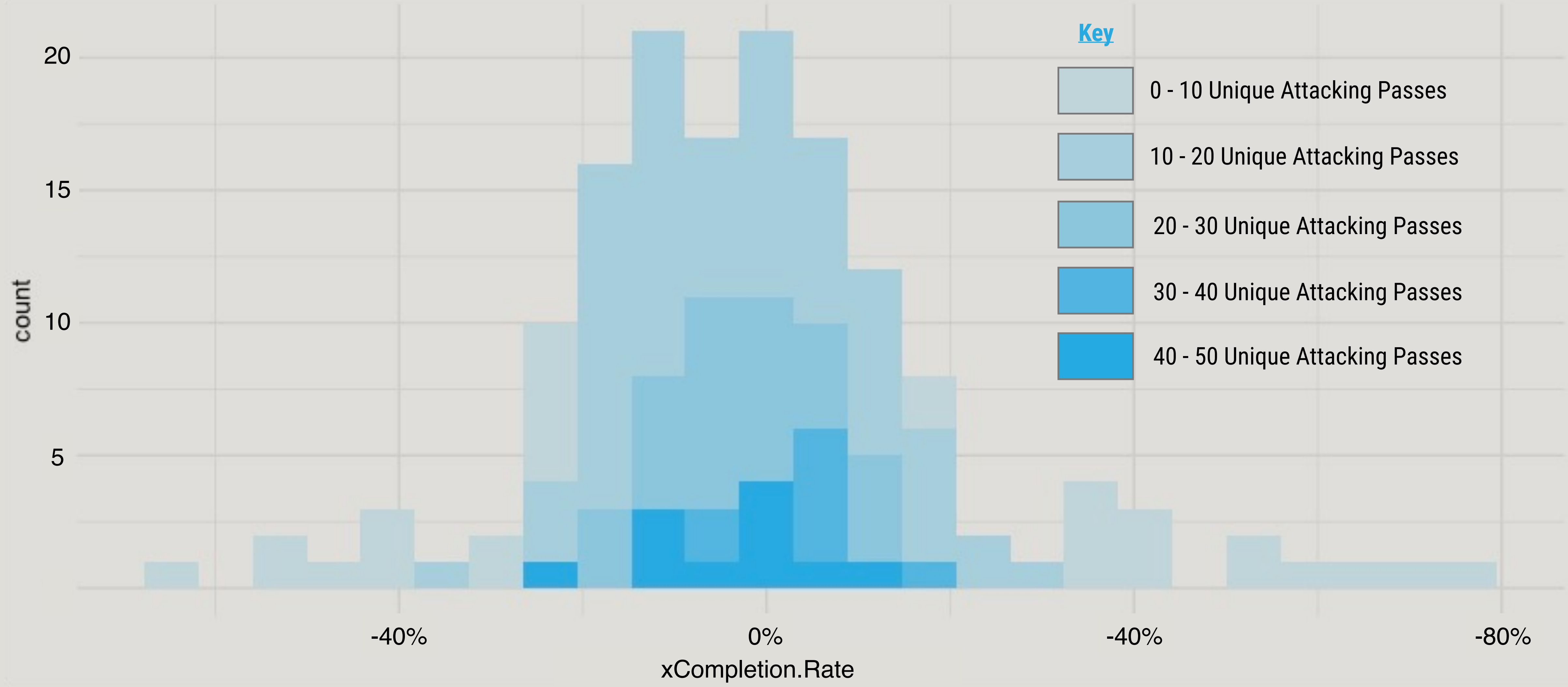
15. REGRESSION TO THE MEAN

There isn't enough detail to the pass completion rates in relation to the uniqueness score.

Adapting the NNS process may give us better success

1. Create a new variable of *success*, completed pass = 1 & incomplete = 0
2. For each pass find the 50 most similar passes
3. Calculate the pass completion rate for those passes as *expected completion*
4. Calculate the xCompletion Difference = *Success - expected completion*
5. Average the xCompletion Difference per player for all UAPs

This could provides a more precise way of assessing if a player passes above expectation. Before shortlisting by xCompletion Difference, it's important to explain why I will filter out all players that have not played above a 20 unique attacking passes. Players with fewer passes can appear extremely low or high, with a higher number of passes most players regress closer to the mean.



16. xCOMPLETION RATES - PER POSITION

DEFENDERS

PLAYER	TEAM	xCOMP.RATE
Thomas Meunier	Belgium	12.2 %
Harry Maguire	England	10.9 %
Jérôme Boateng	Germany	10 %
Raphaël Guerreiro	Portugal	8.6 %
Mário Fernandes	Russia	6.4 %
Sergio Ramos	Spain	6.2 %
Jordi Alba	Spain	1.2 %
Martín Cáceres	Uruguay	0 %
Joshua Kimmich	Germany	-1.1 %
Marcelo	Brazil	-6 %
Benjamin Pavard	France	-6.6 %
Lucas Hernández	France	-7.3 %
Kieran Trippier	England	-8.6 %
Yasir Al Shahrani	Saudi Arabia	-9 %
Hörður Magnússon	Iceland	-10.6 %
Ludwig Augustinsson	Sweden	-11.7 %
Šime Vrsaljko	Croatia	-12.2 %
Henrik Dalsgaard	Denmark	-12.5 %

MIDFIELDERS

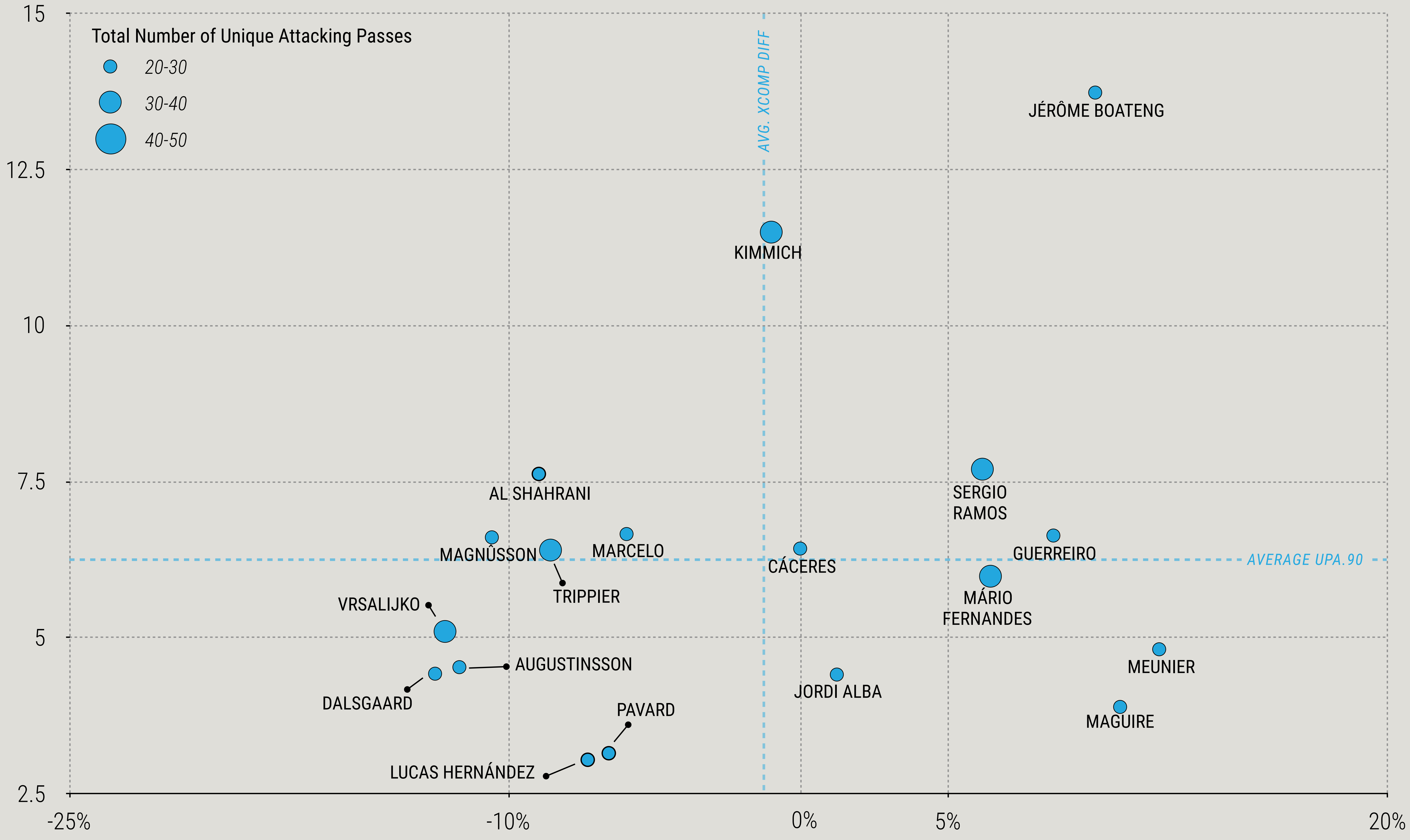
PLAYER	TEAM	xCOMP.RATE
Toni Kroos	Germany	14.9 %
Luka Modrić	Croatia	12.4 %
Nahitan Nández	Uruguay	10.2 %
Lucas Torreira	Uruguay	9 %
William Carvalho	Portugal	8.7 %
Yury Gazinskiy	Russia	7.8 %
Thomas Delaney	Denmark	5.8 %
Sergej Milinković-Savić	Serbia	5.3 %
Dries Mertens	Belgium	5 %
Jesse Lingard	England	4.3 %
Javier Mascherano	Argentina	3.8 %
Philippe Coutinho	Brazil	3.6 %
Granit Xhaka	Switzerland	3.5 %
Christian Eriksen	Denmark	3 %
Andrés Iniesta	Spain	3 %
Ivan Rakitić	Croatia	2.7 %
João Mário	Portugal	1.7 %
Aleksandr Golovin	Russia	1.4 %
Paul Pogba	France	1.3 %
Jordan Henderson	England	0.9 %
Isco	Spain	0.1 %
Viktor Claesson	Netherlands	-0.3 %
Celso Borges Mora	Costa Rica	-1.1 %
Héctor Herrera	Mexico	-1.1 %
Salman Al Faraj	Saudi Arabia	-1.2 %
Sergio Busquets	Spain	-2.2 %
Kevin De Bruyne	Belgium	-5.3 %
N'Golo Kanté	France	-5.8 %
Roman Zobnin	Russa	-6.4 %
Marcelo Brozović	Croatia	-7.5 %
Andrej Kramarić	Croatia	-7.9 %
Dele Alli	England	-16.6 %

FORWARDS

PLAYER	TEAM	xCOMP.RATE
Luis Suárez	Uruguay	5.7 %
Olivier Giroud	France	4.5 %
Neymar	Brazil	3.9 %
Piotr Zieliński	Poland	0.7 %
Eden Hazard	Belgium	0.1 %
Artem Dzyuba	Russia	-1.1 %
Dusan Tadic	Netherlands	-1.4 %
Marcus Berg	Sweden	-4.3 %
Bryan Ruiz	Costa Rica	-5 %
Yussuf Poulsen	Denmark	-6.4 %
Aleksandr Samedov	Russia	-7.1 %
Kylian Mbappé	France	-8 %
Ola Toivonen	Sweden	-8.5 %
Aleksandar Mitrovic	Serbia	-8.9 %
Timo Werner	Germany	-9.2 %
Lionel Messi	Argentina	-9.2 %
Thomas Müller	Germany	-9.2 %
Ivan Perišić	Croatia	-10.3 %
Mario Mandžukić	Croatia	-11.7 %
Harry Kane	England	-14.1 %
Xherdan Shaqiri	Switzerland	-15 %
M'Baye Babacar Niang	Senegal	-18.4 %
Ante Rebić	Croatia	-22.2 %
Antoine Griezmann	France	-22.3 %

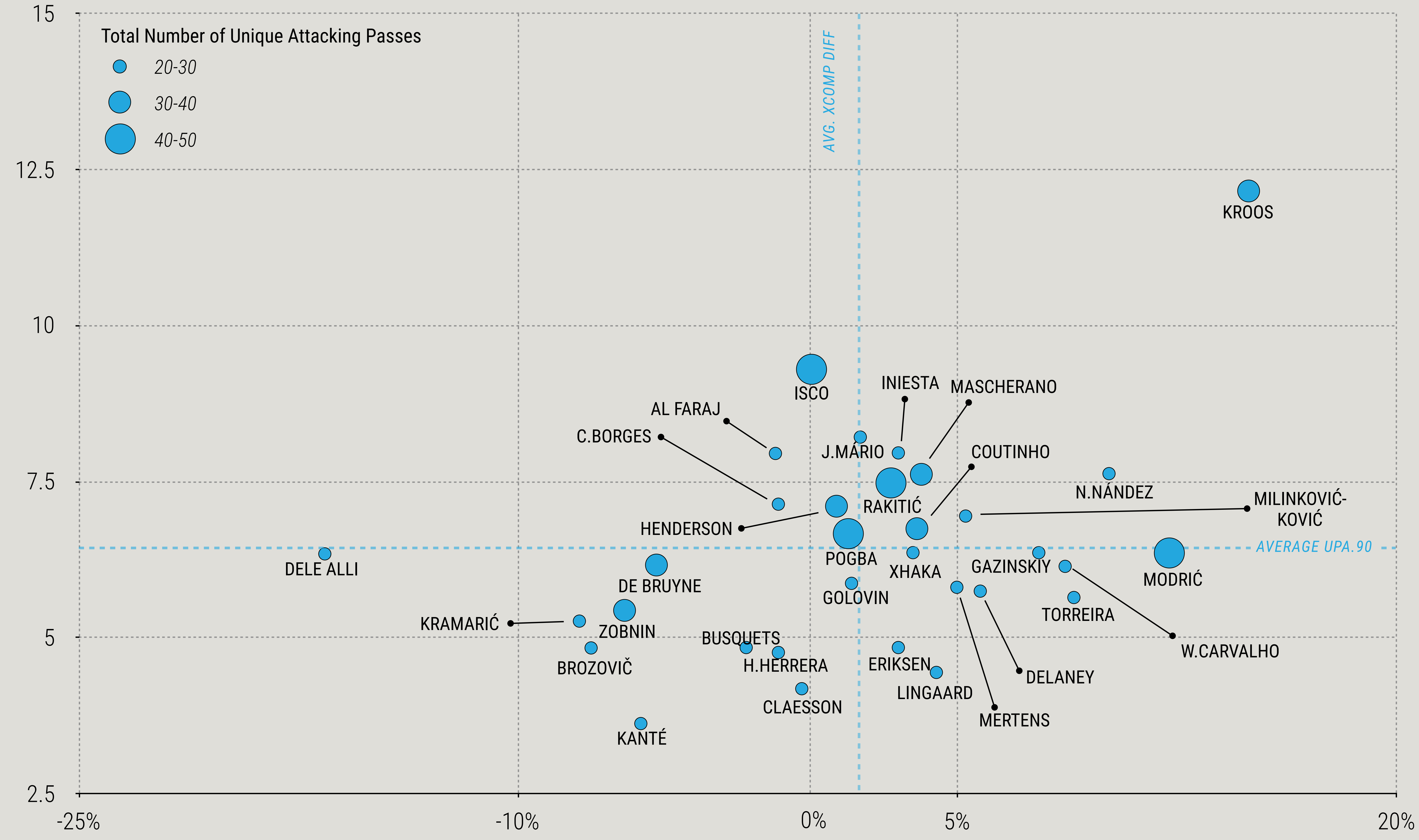
17. PLOTTING THE RESULTS - DEFENDERS

PLAYER	TEAM	xCOMP.RATE
Thomas Meunier	Belgium	12.2 %
Harry Maguire	England	10.9 %
Jérôme Boateng	Germany	10 %
Raphaël Guerreiro	Portugal	8.6 %
Mário Fernandes	Russia	6.4 %
Sergio Ramos	Spain	6.2 %
Jordi Alba	Spain	1.2 %
Martín Cáceres	Uruguay	0 %
Joshua Kimmich	Germany	-1.1 %
Marcelo	Brazil	-6 %
Benjamin Pavard	France	-6.6 %
Lucas Hernández	France	-7.3 %
Kieran Trippier	England	-8.6 %
Yasir Al Shahrani	Saudi Arabia	-9 %
Hörður Magnússon	Iceland	-10.6 %
Ludwig Augustinsson	Sweden	-11.7 %
Šime Vrsaljko	Croatia	-12.2 %
Henrik Dalsgaard	Denmark	-12.5 %



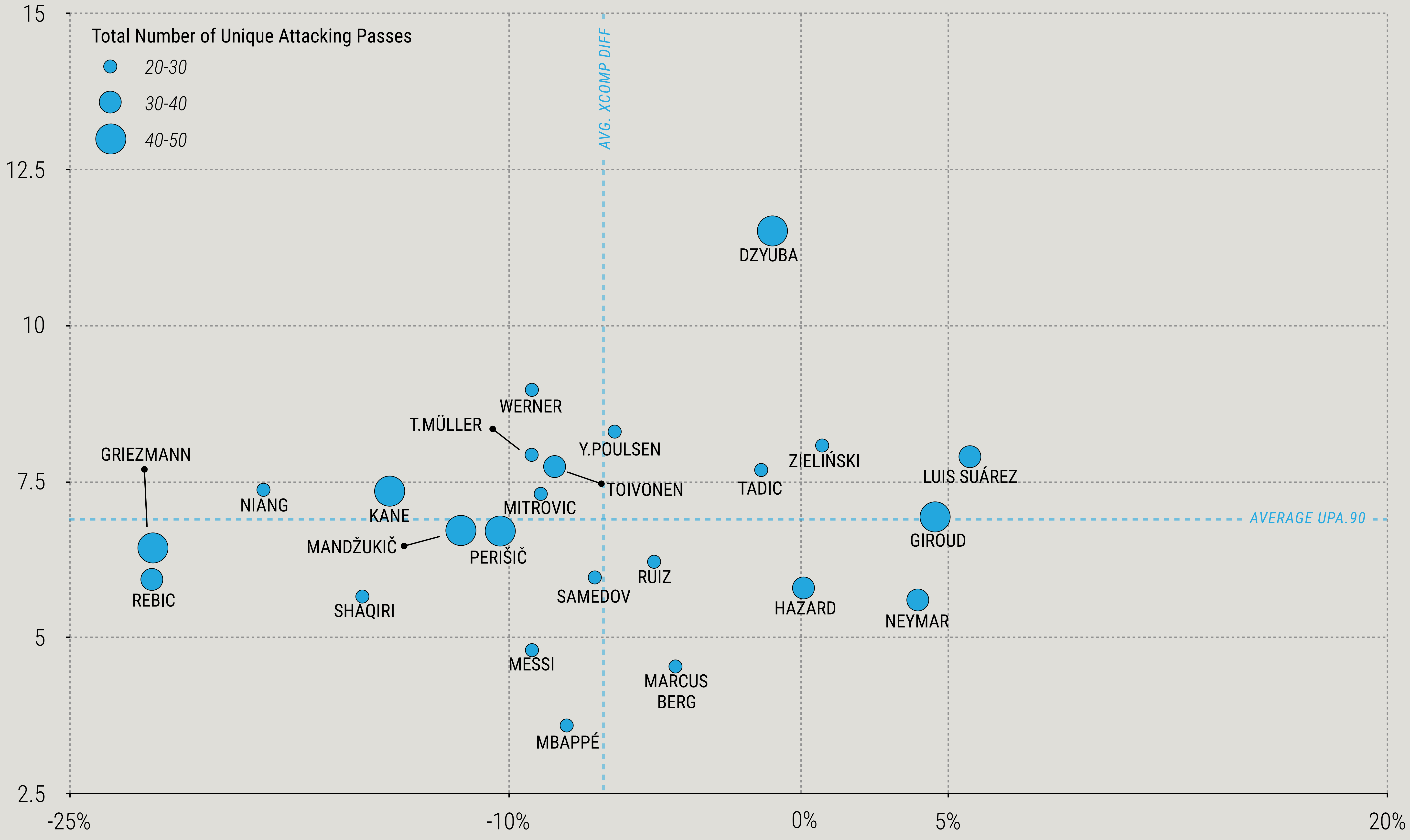
18. PLOTTING THE RESULTS - MIDFIELDERS

PLAYER	TEAM	xCOMP.RATE
Toni Kroos	Germany	14.9 %
Luka Modrić	Croatia	12.4 %
Nahitan Nández	Uruguay	10.2 %
Lucas Torreira	Uruguay	9 %
William Carvalho	Portugal	8.7 %
Yury Gazinskiy	Russia	7.8 %
Thomas Delaney	Denmark	5.8 %
Sergej Milinković-Savić	Serbia	5.3 %
Dries Mertens	Belgium	5 %
Jesse Lingard	England	4.3 %
Javier Mascherano	Argentina	3.8 %
Philippe Coutinho	Brazil	3.6 %
Granit Xhaka	Switzerland	3.5 %
Christian Eriksen	Denmark	3 %
Andrés Iniesta	Spain	3 %
Ivan Rakitić	Croatia	2.7 %
João Mário	Portugal	1.7 %
Aleksandr Golovin	Russia	1.4 %
Paul Pogba	France	1.3 %
Jordan Henderson	England	0.9 %
Isco	Spain	0.1 %
Viktor Claesson	Netherlands	-0.3 %
Celso Borges Mora	Costa Rica	-1.1 %
Héctor Herrera	Mexico	-1.1 %
Salman Al Faraj	Saudi Arabia	-1.2 %
Sergio Busquets	Spain	-2.2 %
Kevin De Bruyne	Belgium	-5.3 %
N'Golo Kanté	France	-5.8 %
Roman Zobnin	Russa	-6.4 %
Marcelo Brozović	Croatia	-7.5 %
Andrej Kramarić	Croatia	-7.9 %
Dele Alli	England	-16.6 %



19. PLOTTING THE RESULTS - MIDFIELDERS

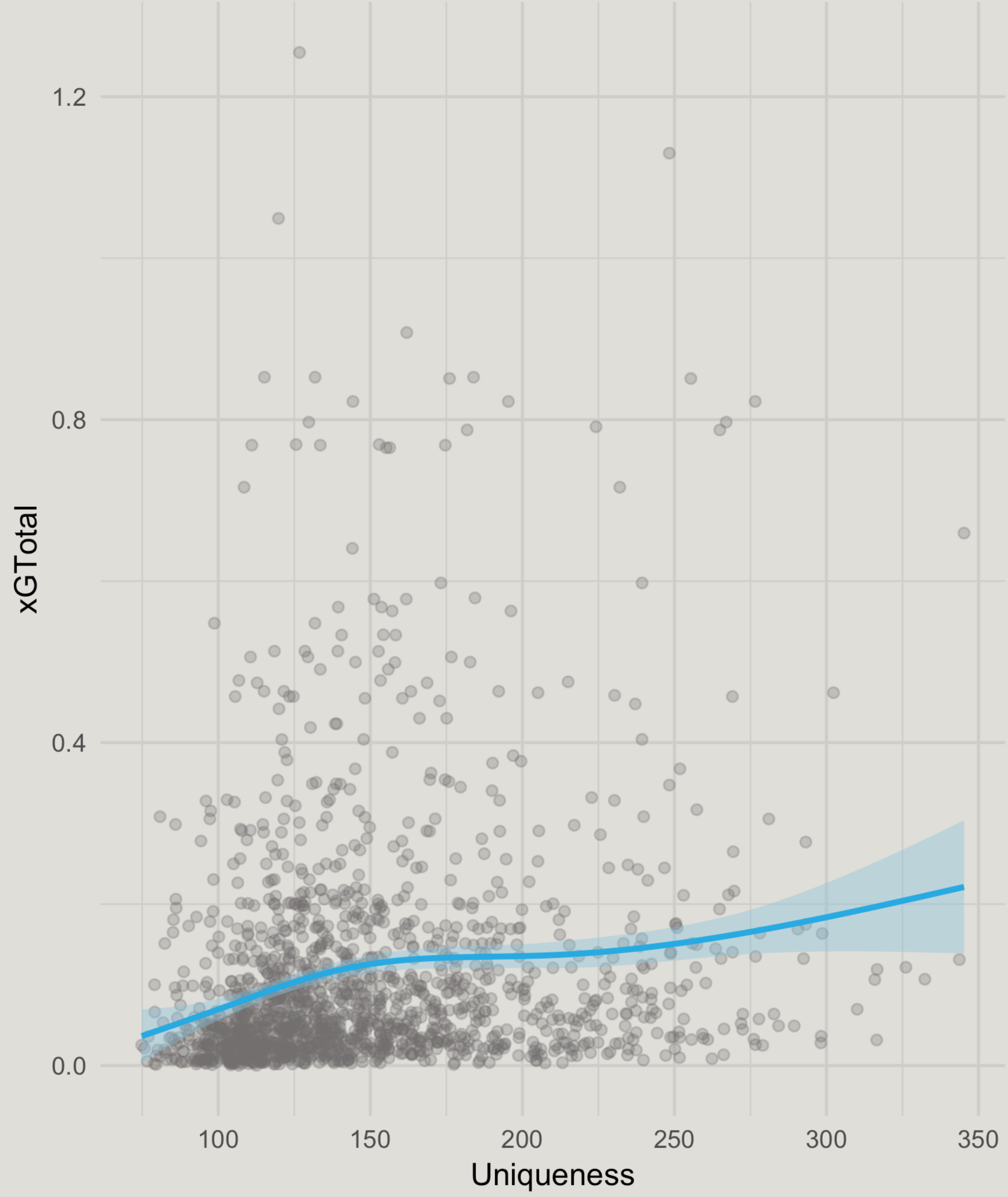
PLAYER	TEAM	xCOMP.RATE
Luis Suárez	Uruguay	5.7 %
Olivier Giroud	France	4.5 %
Neymar	Brazil	3.9 %
Piotr Zieliński	Poland	0.7 %
Eden Hazard	Belgium	0.1 %
Artem Dzyuba	Russia	-1.1 %
Dusan Tadic	Netherlands	-1.4 %
Marcus Berg	Sweden	-4.3 %
Bryan Ruiz	Costa Rica	-5 %
Yussuf Poulsen	Denmark	-6.4 %
Aleksandr Samedov	Russia	-7.1 %
Kylian Mbappé	France	-8 %
Ola Toivonen	Sweden	-8.5 %
Aleksandar Mitrovic	Serbia	-8.9 %
Timo Werner	Germany	-9.2 %
Lionel Messi	Argentina	-9.2 %
Thomas Müller	Germany	-9.2 %
Ivan Perišić	Croatia	-10.3 %
Mario Mandžukić	Croatia	-11.7 %
Harry Kane	England	-14.1 %
Xherdan Shaqiri	Switzerland	-15 %
M'Baye Babacar Niang	Senegal	-18.4 %
Ante Rebić	Croatia	-22.2 %
Antoine Griezmann	France	-22.3 %



20. xG & Uniqueness

The idea in my head of a ‘unique’ pass at the start of this investigation was a creative, defence splitting and shot creating pass.

Let’s analyse the relationship between uniqueness and expected goal value of resulting shots. Find all shots within 5 seconds of a unique attacking pass and sum the expected goals value.



There does seem to be a mild positive correlation. However, the relationship was not statistically significant. With more data these conclusions could be tested with greater clarity.

21. FINAL THOUGHTS

1. The Uniqueness Passing Model seems to help with shortlisting and player identification for further assessment. However, the same could be said about passes in the final third per 90 (and there is minimal computational requirements!). It 'works' but it's pretty clunky, & sellotape together

2. I don't have enough data to thoroughly test if UAPs is replicable and predictive to shots, expected goals or goals themselves. I would be interested to investigate this further if (when) Statsbomb release more data to play with.

3. I hope the working example of the NNS will prompt further investigation into how it can be used more extensively within public football analytics.

4. The Statsbomb dataset is amazing, well designed and a real gift to the community. We are so lucky to have access to it! So utilise it to death!