

# Mathematical modelling of football

David Sumpter

Uppsala University & Hammarby IF



UPPSALA  
UNIVERSITET



STIFTELSEN för STRATEGISK FORSKNING

# Structure today

10:15-11:00 Information about the course

11:15-12:00 Visualising actions on the pitch

Please ask questions in the **comments** section where possible. That will give me time to answer them.

I will take a reasonably slow pace and interact along the way.

# A little bit of history....



**Javier Fernandez**  
Soccer Data  
Scientist  
FC Barcelona



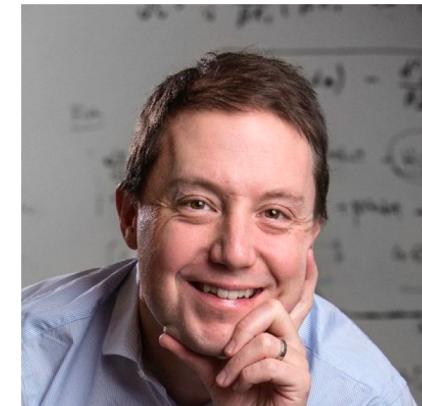
**Suds Gopaladesikan**  
SL Benfica Sports  
Data Science



**Fran Peralta**  
Football Data  
Scientist  
Hammarby IF



**Pascal Bauer**  
Soccer Data-Scientist  
German National  
Team (DFB)



**David Sumpter**  
Soccermathematician  
Uppsala/Hammarby

## Q&A

Limitations?

HOW TO ACCOUNT MORE THE SHORT TERM ACTION?  
IT'S DOABLE WITH THE FRAMEWORK

Possession BASED TEAMS  
GET MORE POINTS?  
NOT NECESSARY GETS HIGHER RATINGS

BETTER THAN HUMAN SCOUT?  
VS

KNOW THE CHEMISTRY between them!  
PLAYERS PAIRS

A LOT TO LEARN FROM EVENT AND TRACKING DATA

PAPERS BLOGS  
Check YOUTUBE VIDEO FOR complete list!

TUTORIALS

online coming soon...

NOT HAVING THE BENEFITS OF TRACKING DATA

GOAL KEEPER ACTIONS?

A LOT IS NOT CAPTURED BY EVENT DATA

XGoal MODEL FOR THEM



# FRIENDS OF TRACKING

HOW TO VALUE ACTIONS USING EVENT DATA

SciSPORTS

## VAEP RATINGS

IDENTIFY MOST VALUABLE PLAYERS

OUR FRAMEWORK!

ACTIONS RATINGS ANSWER A LOT OF QUESTIONS

Hidden Gems

VALUE +



LOTTE BRANSSEN

TRADITIONAL STATISTICS FAIL TO CONSIDER CONTEXT



SARAH RUDD (2011)

EVENT DATA

(WHAT HAPPENS BEFORE)  
(Messi receive the BALL)  
PRE ACTION  
POST  
EXAMPLE  
Difference Between GAME STATES

STYLE ACTIONS

## PLAYER

Help SCOUTING

<1% balls actions are SHOTS

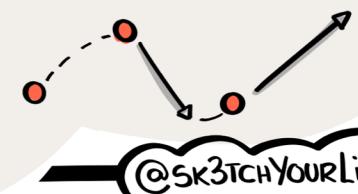
## FOOTBALL

Different FLAVORS

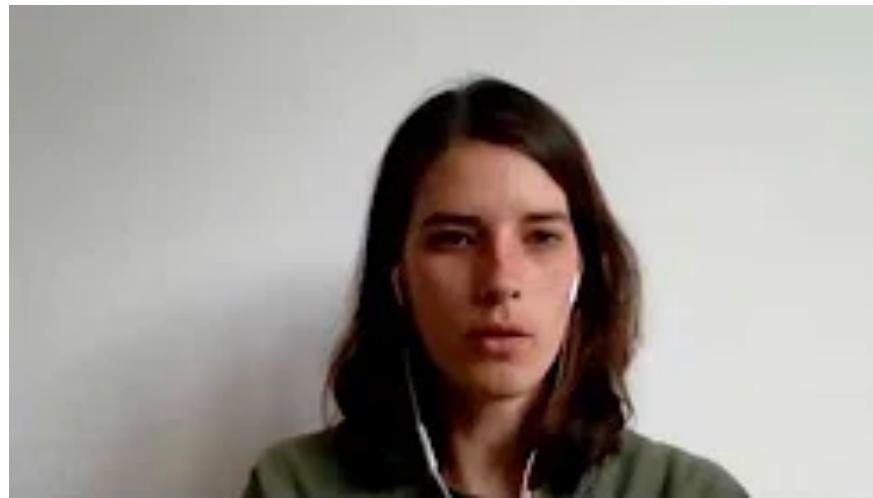
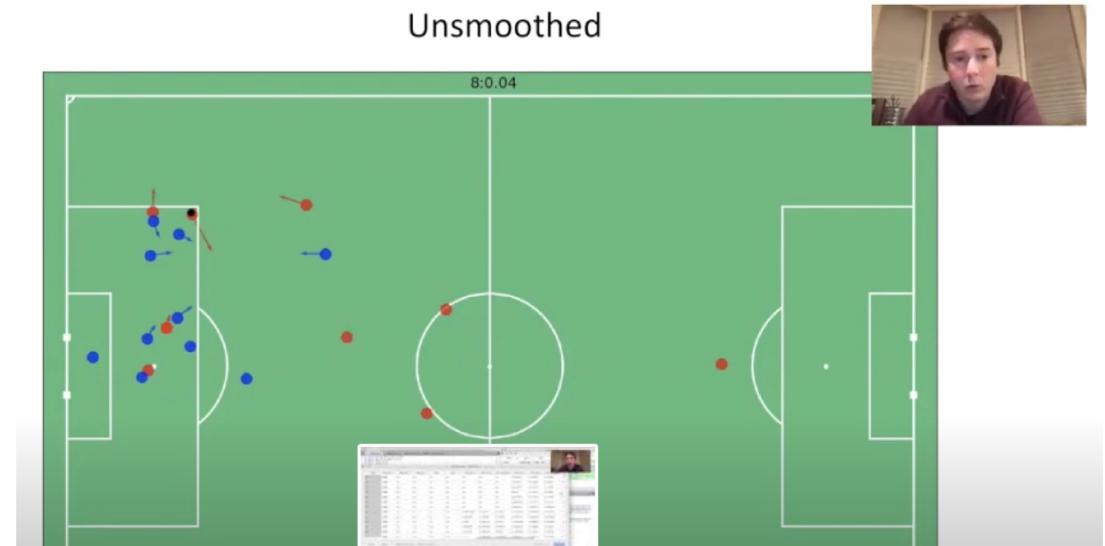
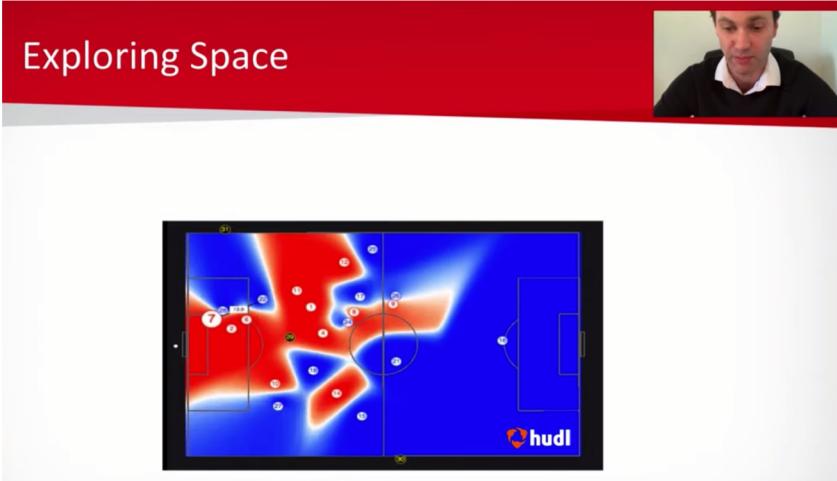
## FRAMEWORKS

THAT TAKE PLAYER ACTION into account

MORE THAN ONE FRAMEWORK [differs in several aspects!]



@SKETCHYOURLIFE



# The course at Uppsala University

Mathematical Modelling of Football

[Jump to today](#)

## Mathematical Modelling of Football

### First Lecture

The first lecture is 10:15 on September 1st and will be be streamed live over Youtube here:



<https://uppsala.instructure.com/courses/28112>

A little bit on the registration process...

# A Slack group for those who couldn't register

Mathematical Modelli... 

- @ Threads
- >All DMs
- @ Mentions & reactions
- : More
- ▼ Channels +
  - # admin-questions
  - # general
  - # mini-challenge-plotting-ac...
  - # other-subjects
  - # study-group
  - # update**
- ▼ Direct messages +
  - ...

#update ☆  
Add a topic

Saturday, August 29th

  
UPPSALA

4 2 1

It will remain on Youtube for those who are still in bed!

Rodrigo Salvador 9:10 AM  
was added to #update by david sumpter.

Today

david sumpter 9:12 AM  
First lecture in an hour!

# Machine Learning and Statistics - Master's Programme in Data Science 2020/2021



- 120 credits
- Uppsala
- Campus 100%
- [Programme Syllabus and Outline](#)

**Application →**

**OVERVIEW**

PROGRAMME

CAREER

ADMISSIONS

CONTACT AND FURTHER RESOURCES

As a data scientist, you will learn to extract valuable insight from one of the most important resources today - data. Use the latest machine learning methods to turn large amounts of information into big-picture knowledge.

**Register your  
interest**



**UPPSALA  
UNIVERSITET**

# Form of Teaching

- The main form of learning is through online videos and doing things yourself.
- Tuesdays 10:15 – 12:00 will be more of a lecture form. I will bring the information in the videos together and give the ‘bigger picture’. I will also answer questions.
- Thursdays 14:00 – 17:00 will be Zoom tutorials in groups of 10 students. You will be randomly assigned groups (Tutorial teams).
- Tutorial teams (and Slack Group) should support each other. Help each other. But, for the first exercises, you should complete and submit your own work.
- You will work together in the tutorial teams on the final projects.

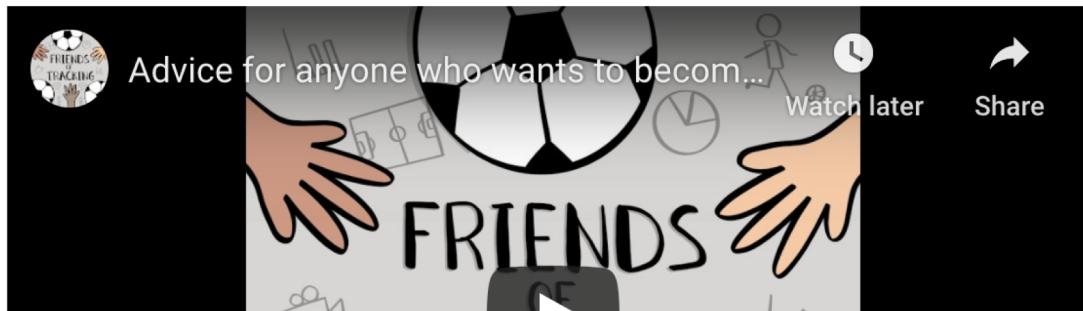
# Watch the getting started videos now if you haven't already

## 0. Getting started

We recommend you start working through this material *before* the course starts on August 31st. We've included some basic Python and R code snippets to help you get started with coding when the course starts.

### What does a football data scientist do?

In April 2020, David interviewed four football data scientists (Pascal Bauer, Deutscher Fußball-Bund; Gopaladesikan, Benfica and Fran Peralta, Hammarby) about their role at their clubs and national team working within clubs and what they do. This discussion offers a great starting point to finding out what a football data scientist does.



# And watch the plotting actions videos too.

## 1. Visualising actions

This week we will work with event data. Here we assume you have watched the [setting up for the course video](#) at the bottom of 'wee have set up an environment where you can program in Python. Please work through this webpage, doing the exercises as you go. It is you can work with sections 1.1 - 1.3 before the first lecture on Tuesday 1st September at 10:15.

### 1.1 Handling Event Data

In this lecture I go through the basics of loading in event data from Statsbomb.



# Start with the mini-challenge

## 1.3 Mini-challenge: plotting actions

1, Think of a player who you enjoyed watching at the recent Men's or Women's World Cups.

2, What actions did they perform that were important and why?

3, Plot the actions and describe how the data supports or contradicts your own analysis.

4, Write a short text using at most two figures that illustrate your point.

Go in to the Discussion group for this mini-challenge and post your answers. You are very welcome to use methods discussed in later lectures here.

Deadline is the evening of Wednesday 9th, before the tutorial in course week 2.

In the discussion forum or Slack:

- Chat with each other.
- Give feedback.
- Resubmit a better version.
- Work and learn.

# Same process each week

- Start watching material before the *Tuesday* lecture.
- Make sure you have done analysis/programming before the *Thursday* session.
- Chat with each other both in your tutorial group and in other groups.

# Timetable

## Section 1: Event data

31<sup>st</sup> August: [Visualising actions](#)

7<sup>th</sup> September: Expected Goals model

14<sup>th</sup> September: Simulation of matches

20th September: Deadline for hand-in 1.

## Section 2: Tracking data

21<sup>st</sup> September: Player movements on the pitch.

28<sup>th</sup> September: Pitch control.

4<sup>th</sup> October: Deadline for hand-in 2.

## Section 3: Applications and Advanced analytics.

5<sup>th</sup> October: Scouting and evaluating actions.

12<sup>th</sup> October: Expected Possession Value.

17<sup>th</sup> October: Incorporating fitness data.

24<sup>th</sup> October: Work on final project.

30<sup>th</sup> October: Final deadline for hand-in 3.

# Navigate inside Canvas

# Navigate inside Slack

# Lecture 1

# Visualising Actions on the Pitch

Mathematical modelling of football

BBC 18:49 GER 0-0 ITA

VIDEORAM

# Pass sequence

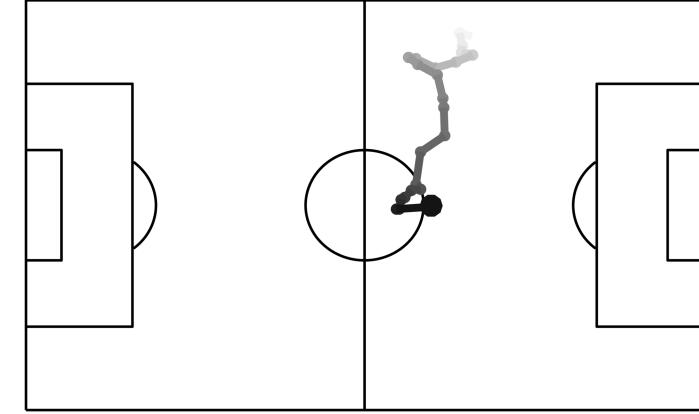
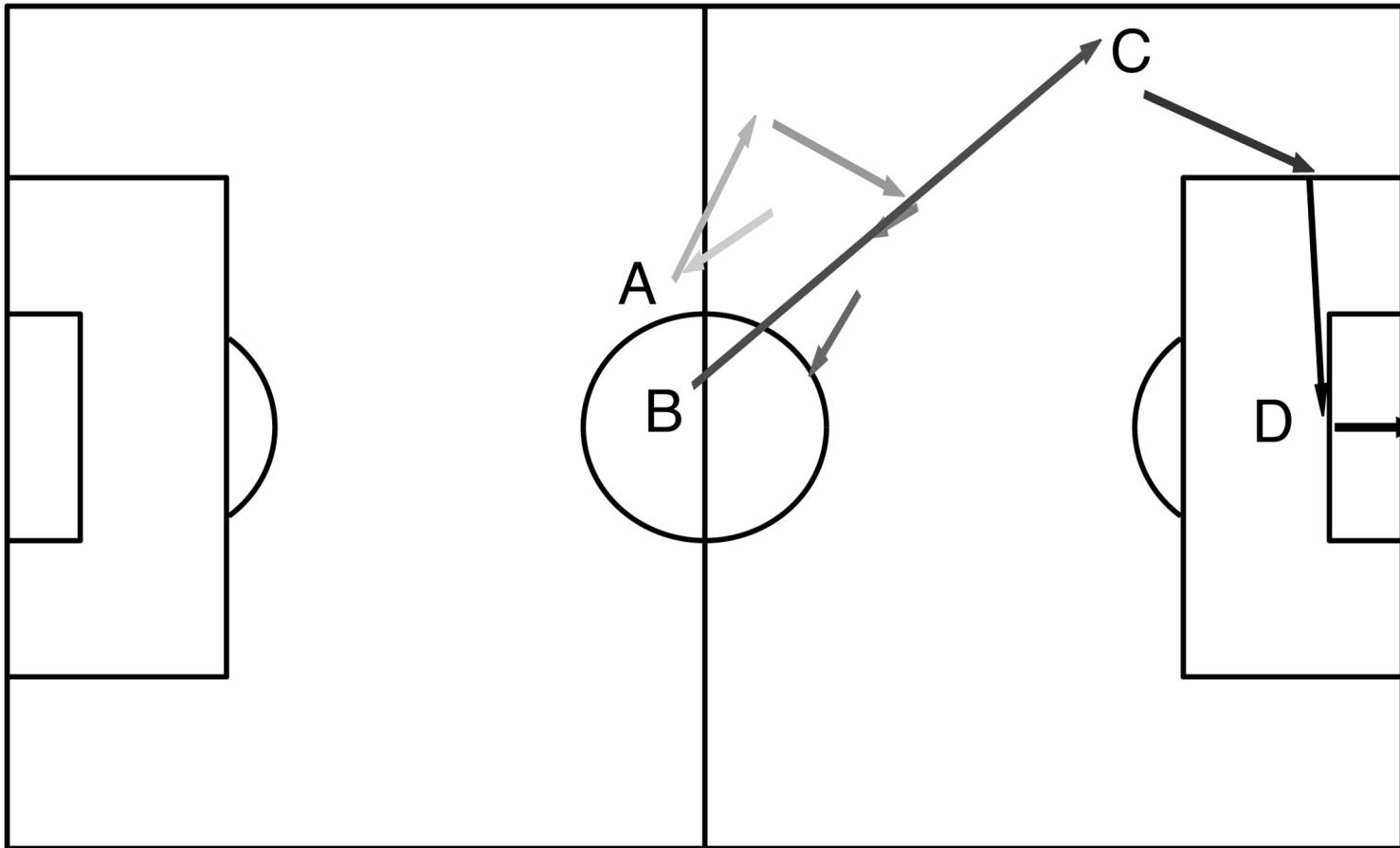
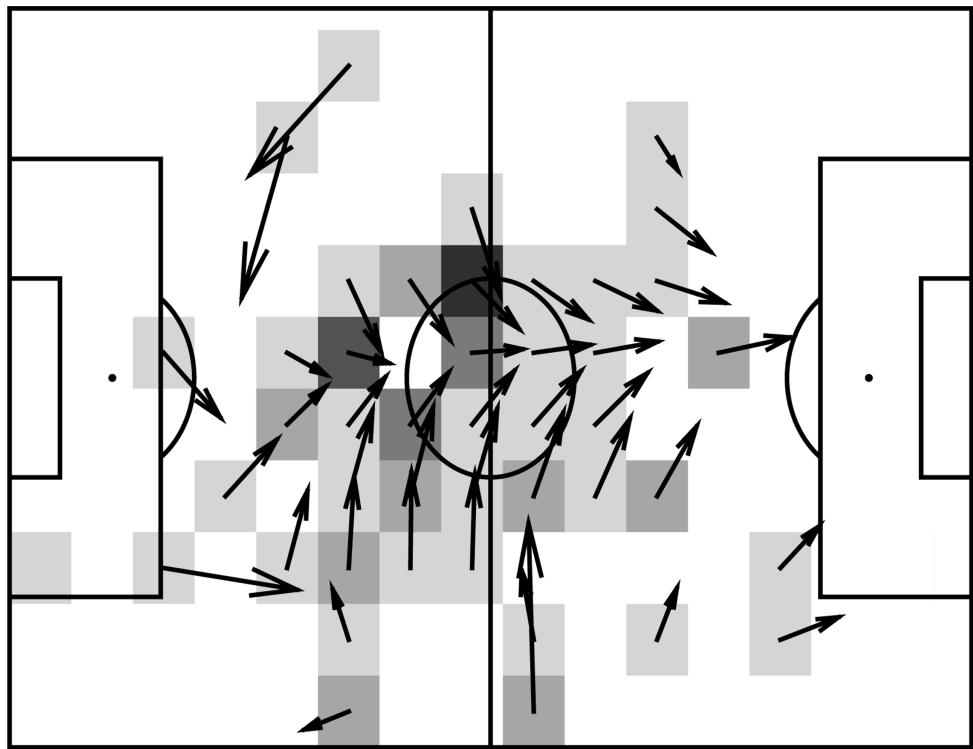
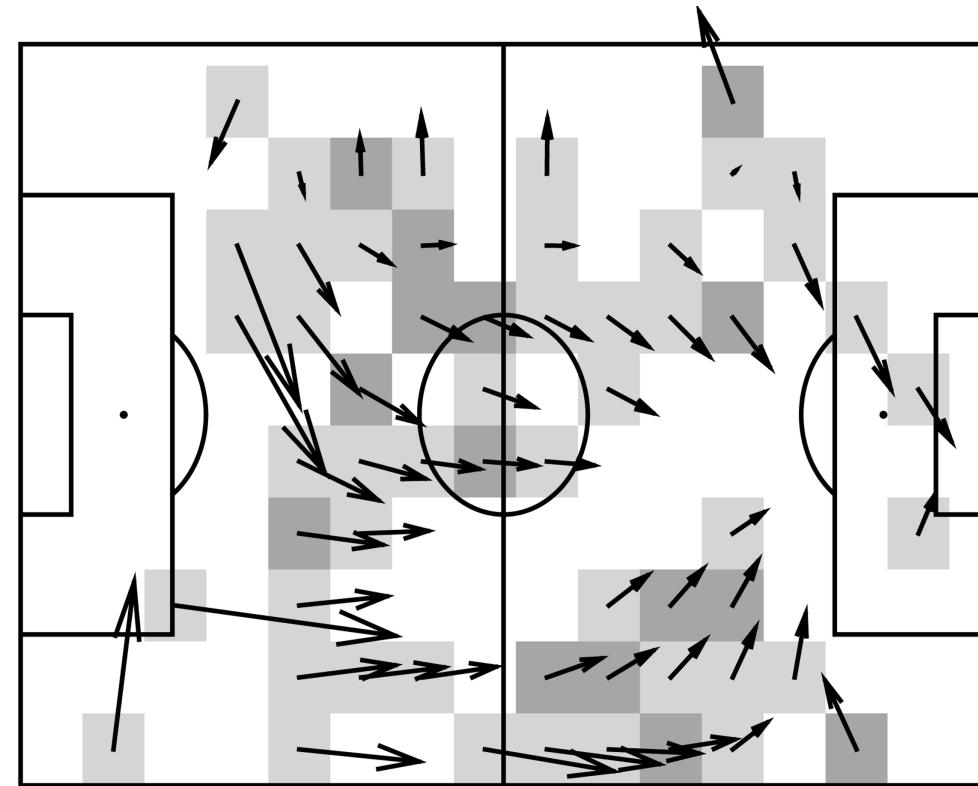


Figure 3.6 Passes leading up to Italy's first goal against Germany in Euro 2012, showing passes made by Italy (arrows in top panel) and Mesut Özil's movement while chasing the ball (bottom panel). Darker shading indicates more recent events in time. Letters indicate events: (A) Pirlo's first pass; (B) Pirlo's second pass; (C) Chiellini receives the ball; (D) Balotelli scores.



**Figure 3.7** Pirlo's passing during Italy's semi-final against Germany in Euro 2012. The intensity of the shading at each particular position on the pitch indicates how often Pirlo made a pass from this position. The arrows are a statistical fit to all 66 passes he made during the match. The average direction of passes is indicated by the direction of the arrow. The average length of the arrow is proportional to the average length of a pass.

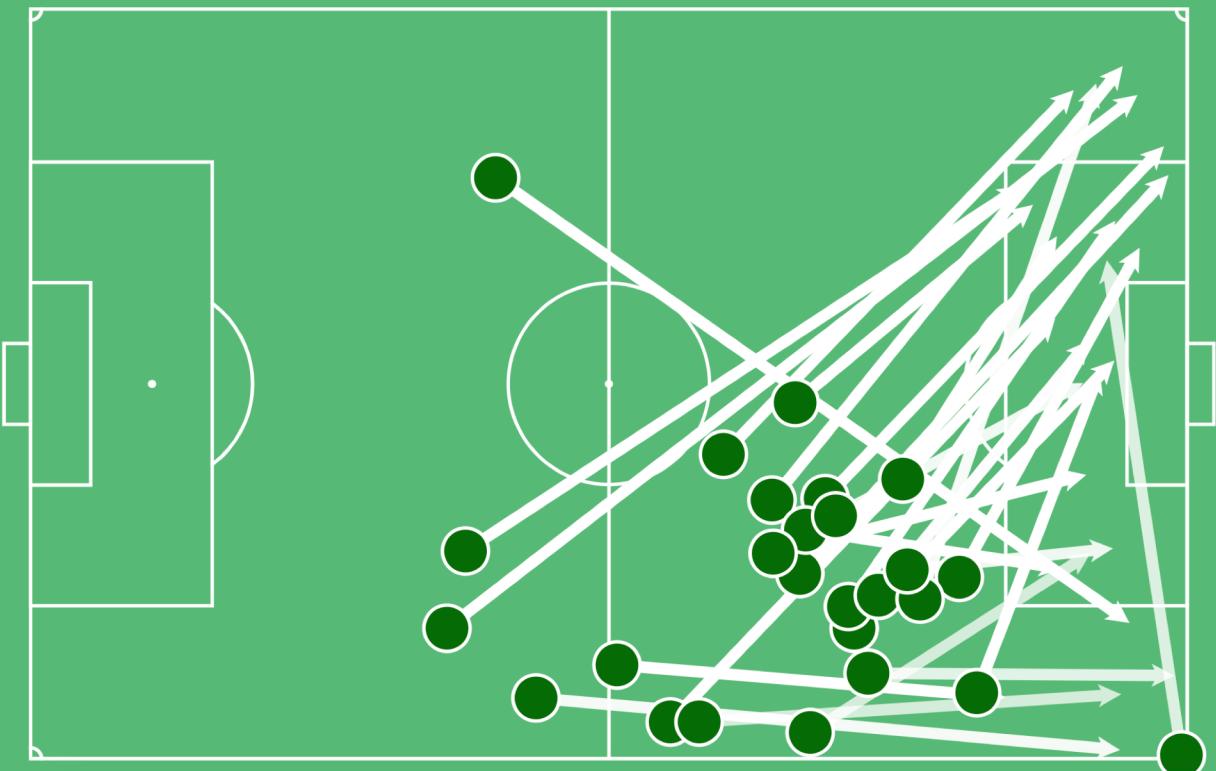


**Figure 3.8** Schweinsteiger's passing during Germany's semi-final against Italy in Euro 2012. See Figure 3.7 for details of how this plot was created.

**3**

## TRENT ALEXANDER-ARNOLD

ATTACK



6,233 points

10 matches • 753 mins

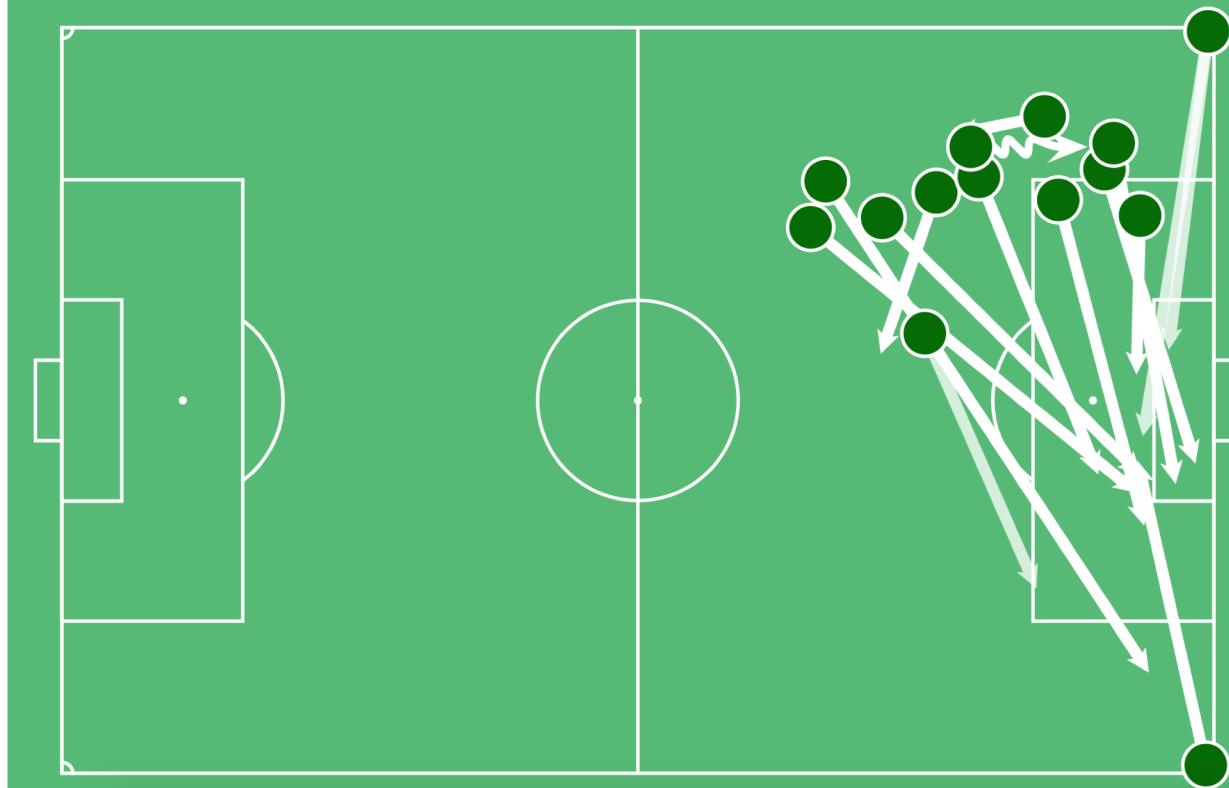
**2**

## ANDREW ROBERTSON

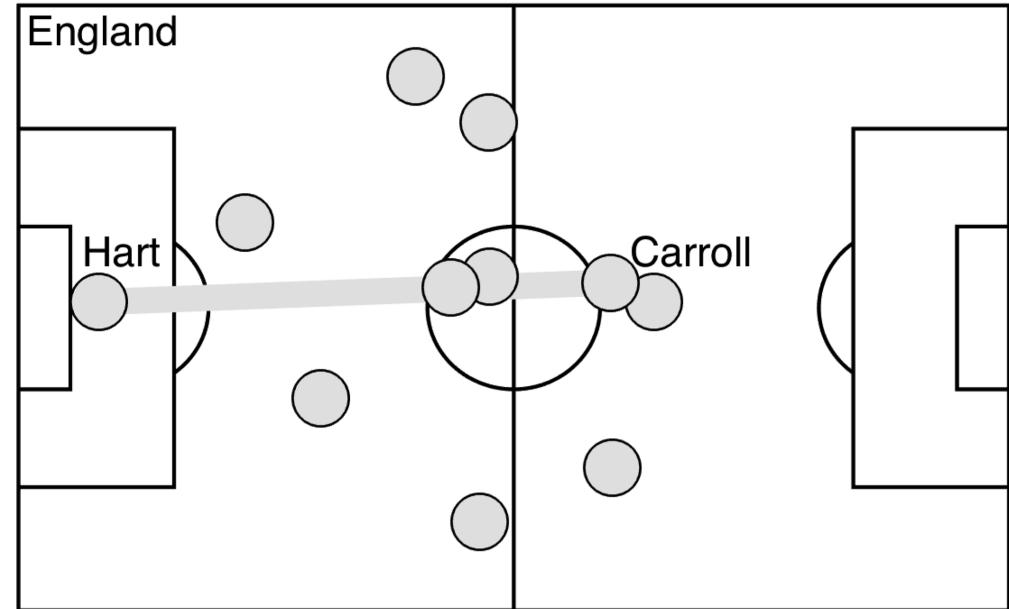
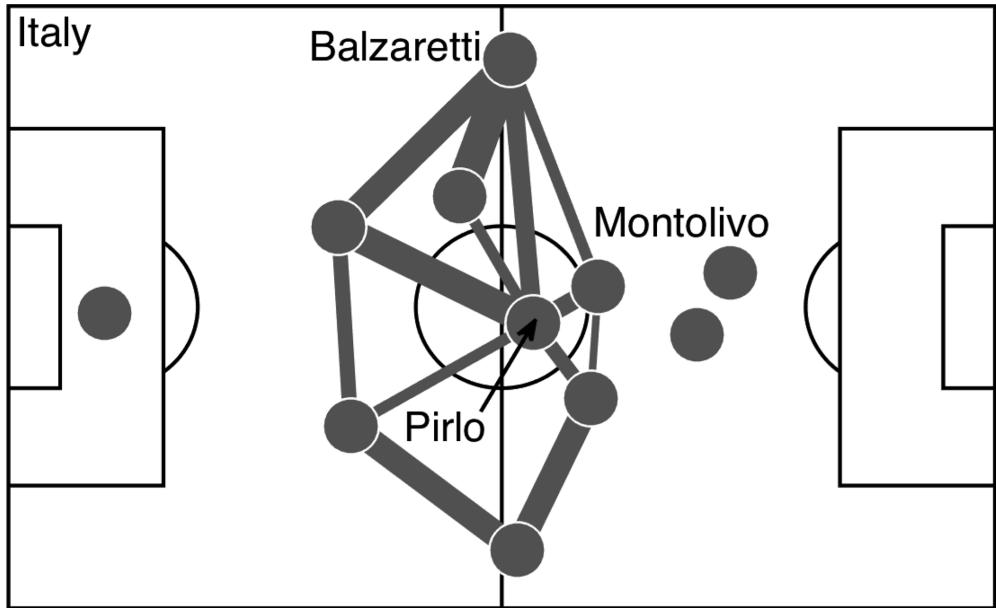
ATTACK

6,607 points

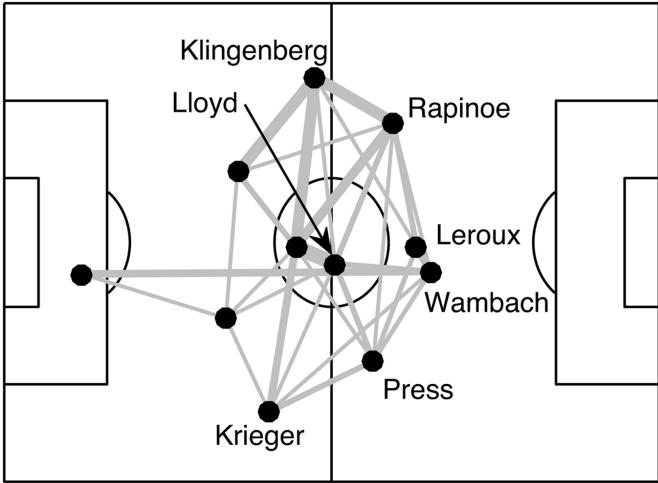
10 matches • 702 mins



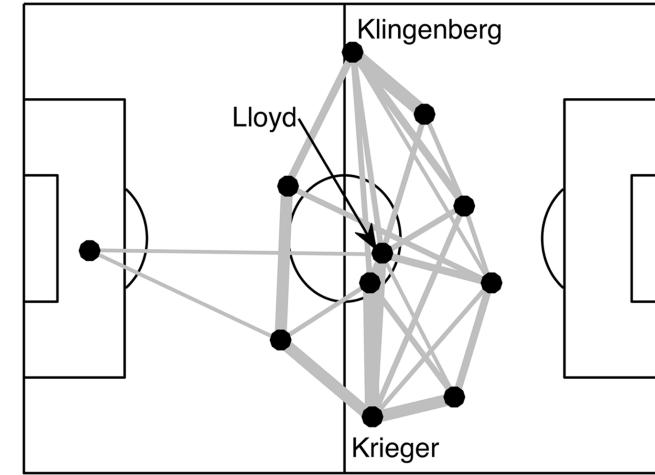
<https://www.dropbox.com/s/e177q1r524pxfeu/Screenshot%202020-08-28%20at%2011.38.13.png?dl=0>



**Figure 7.1** Passing networks for Italy (top) and England (bottom) in the Euro 2012 quarter-finals, showing links between pairs of players that completed 13 or more passes between them. Thicker lines indicate a higher number of passes. Both teams are attacking left to right.

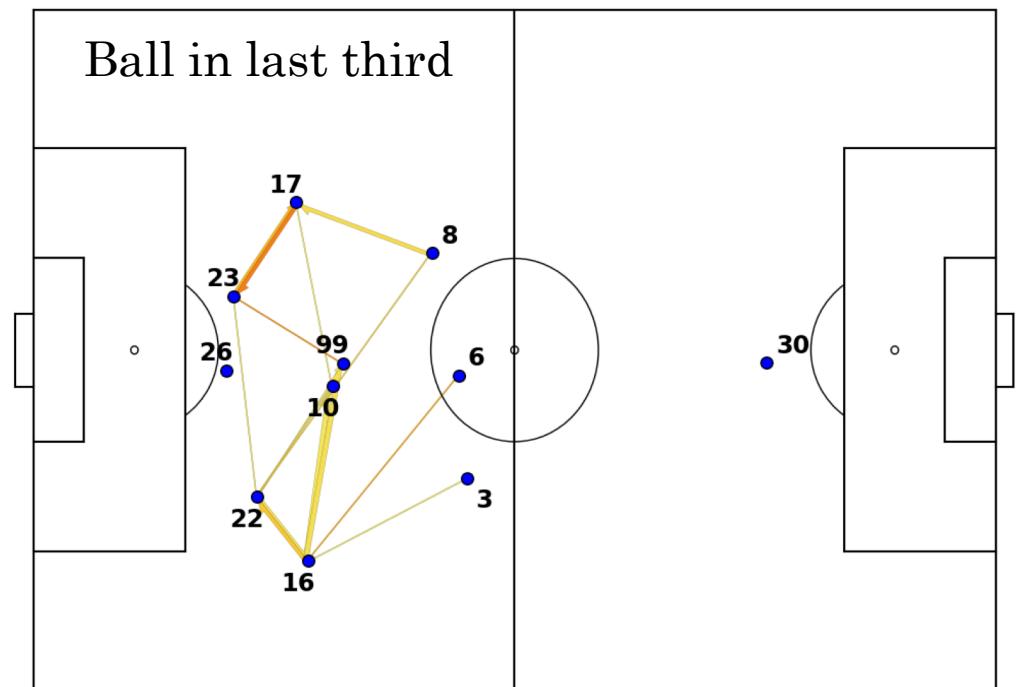
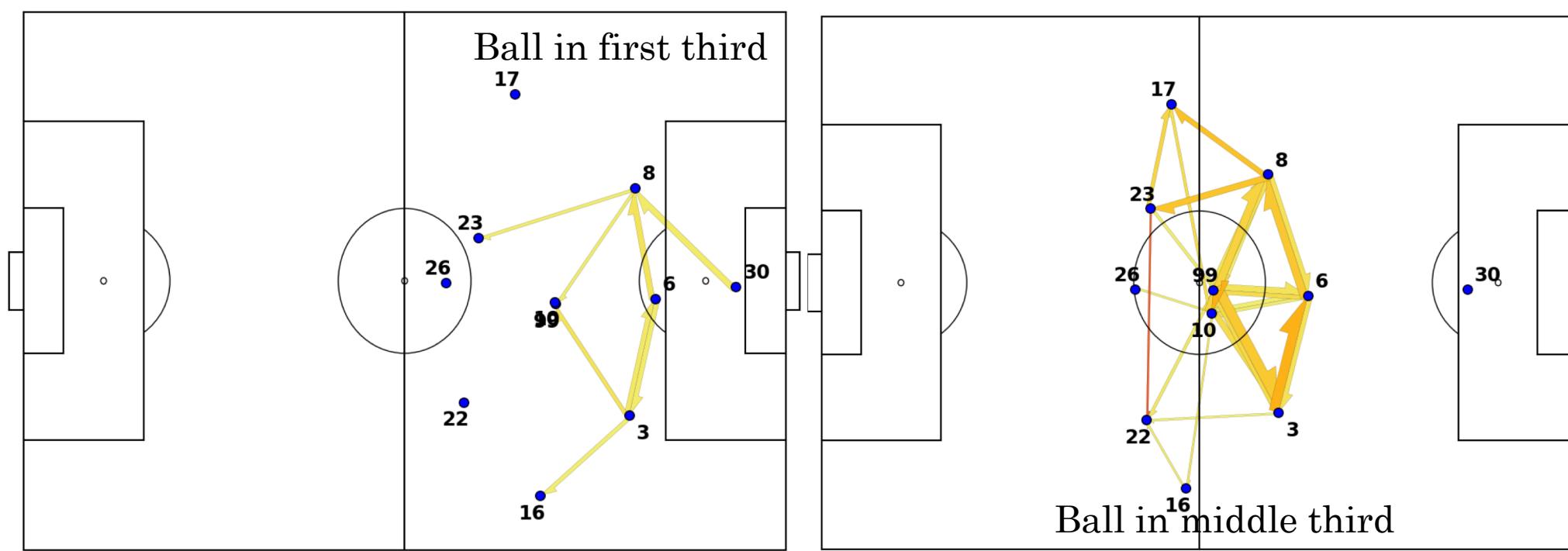


**Figure 7.3** Passing network for the US National team playing against Australia in the Women’s World Cup 2015, showing links between players that completed five or more passes between them. Thicker lines indicate a higher number of passes. Original data analysis by Devin Pleuler.



**Figure 7.4** Passing network for the US National team playing against China in the Women’s World Cup 2015. For details see Figure 7.3. Original data analysis by Devin Pleuler.

# FIRST HALF

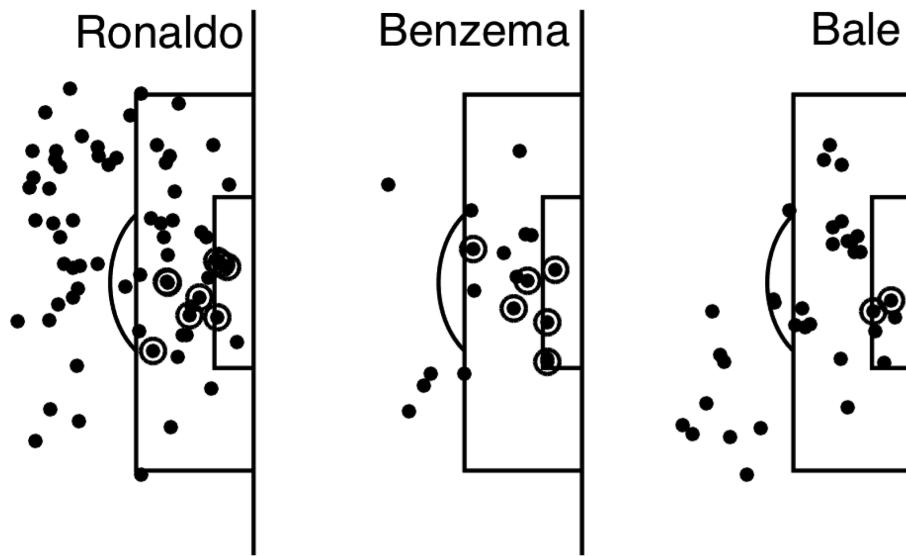


## IK Sirius FK

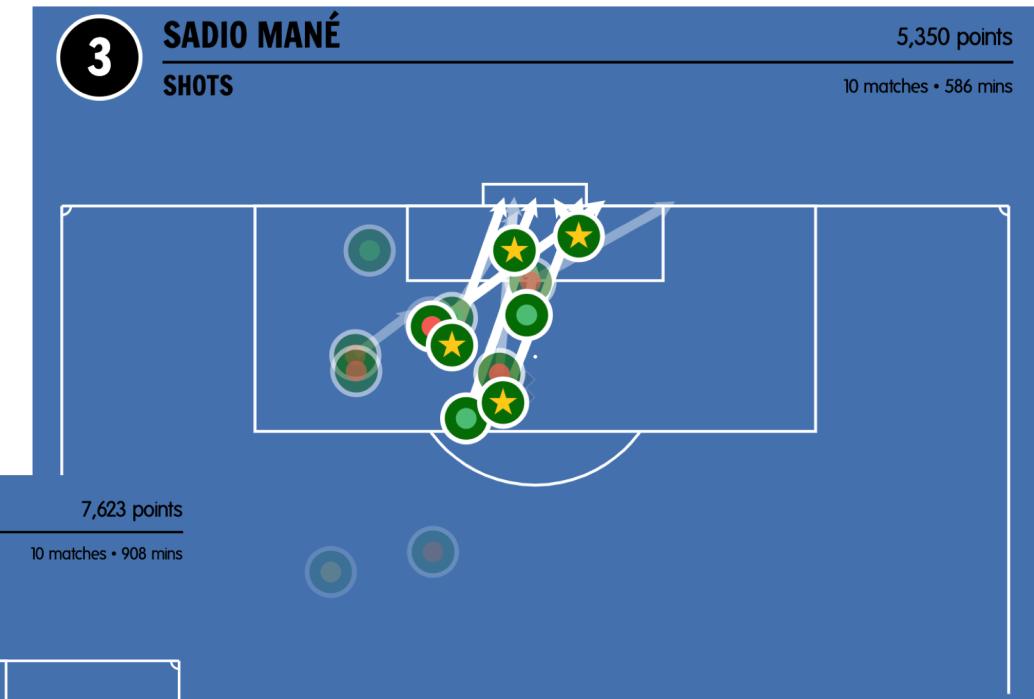
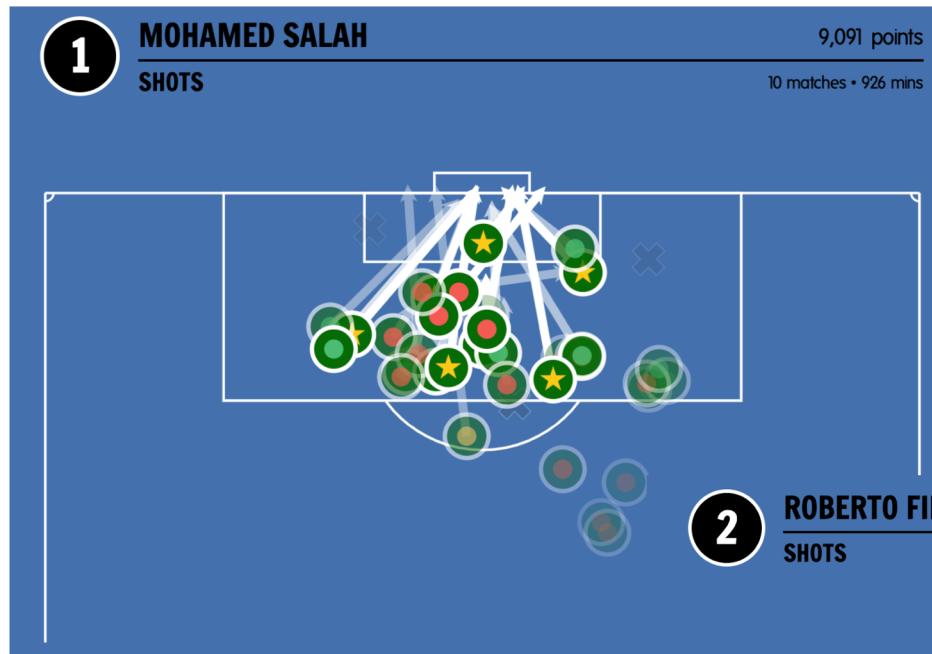
- 30 - Jonathan Viscosi
- 3 - Karl Larson
- 8 - Tim Björkström
- 6 - Hjalmar Ekdal
- 17 - Adam Ståhl
- 99 - Nahom Girmai Netabay
- 10 - Elias Andersson
- 16 - Axel Björnström
- 23 - Yukiya Sugita
- 26 - Mohammed Khalid Saeid
- 22 - Stefano Holmquist Vecchia



# Shot maps

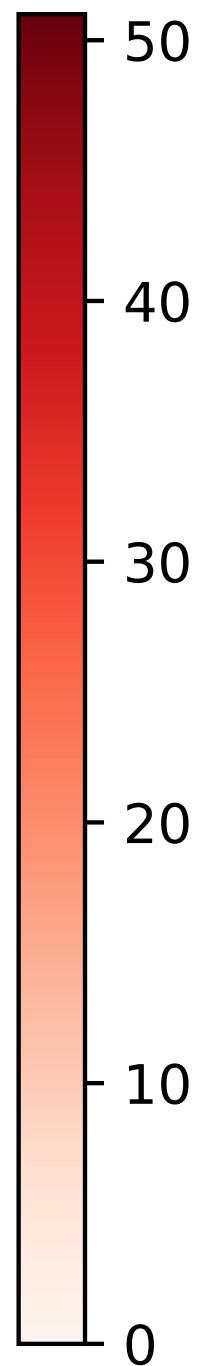
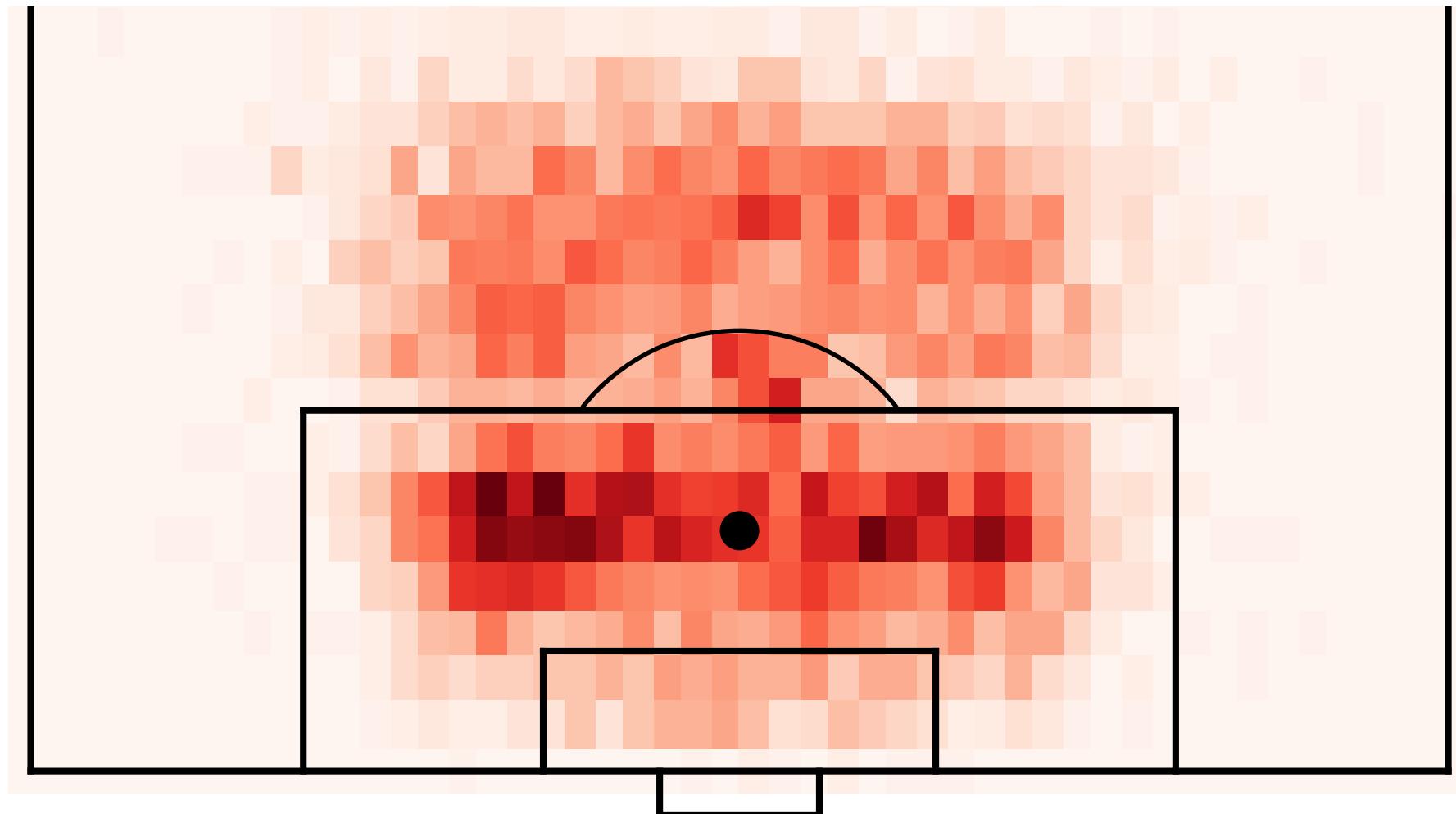


**Figure 7.11** Shots made by Cristiano Ronaldo, Karim Benzema and Gareth Bale during the Champions League season 2014/15. The dots represent the positions the players took a shot from, and the rings the positions from which they scored. Data provided by Opta.

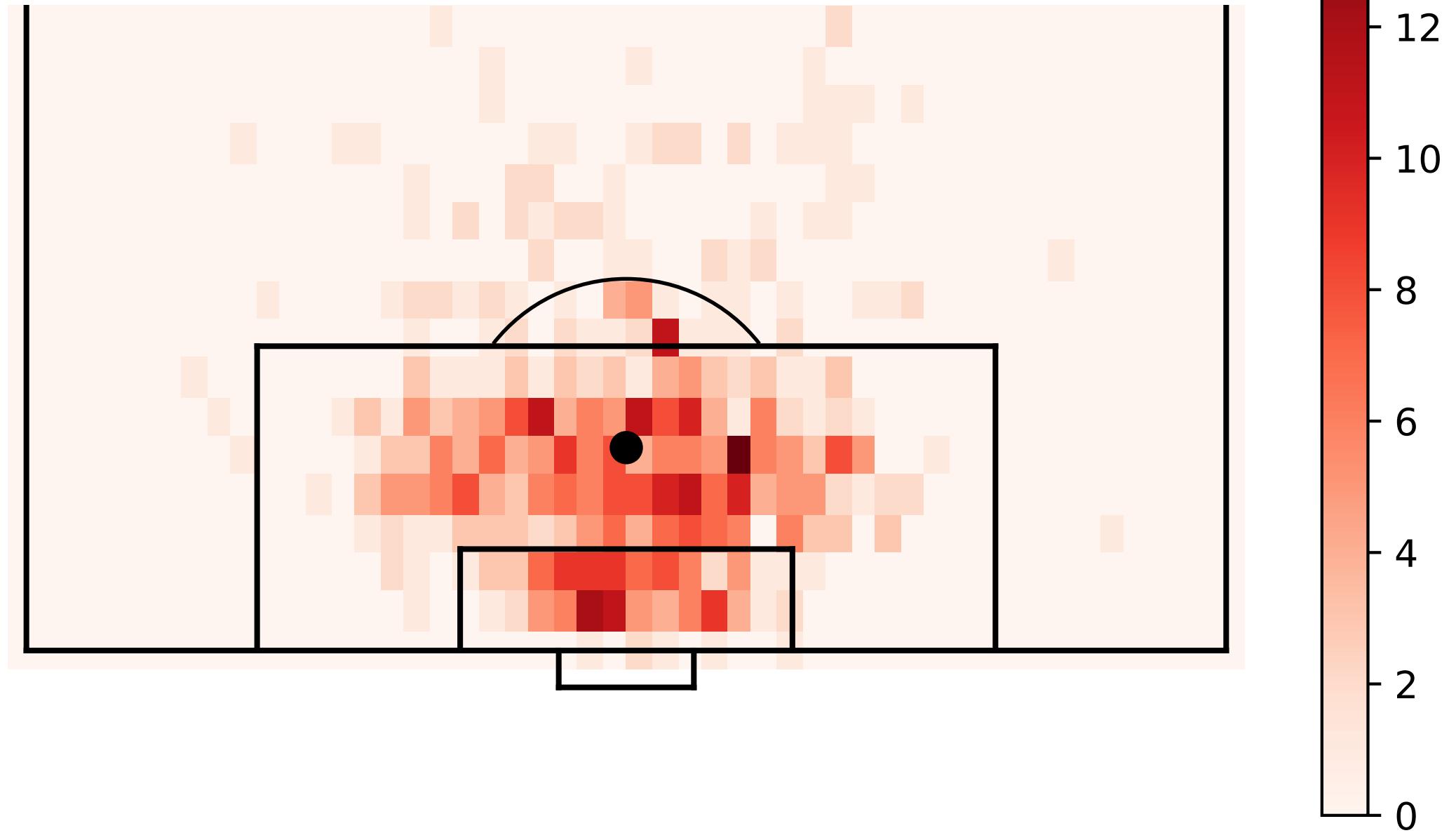


<https://twelve.football/analytics?analysisId=334ff46fa6794015b243b67e4ec2a31a&viewType=shot>

# Number of shots



# Number of goals



# Touch heat maps

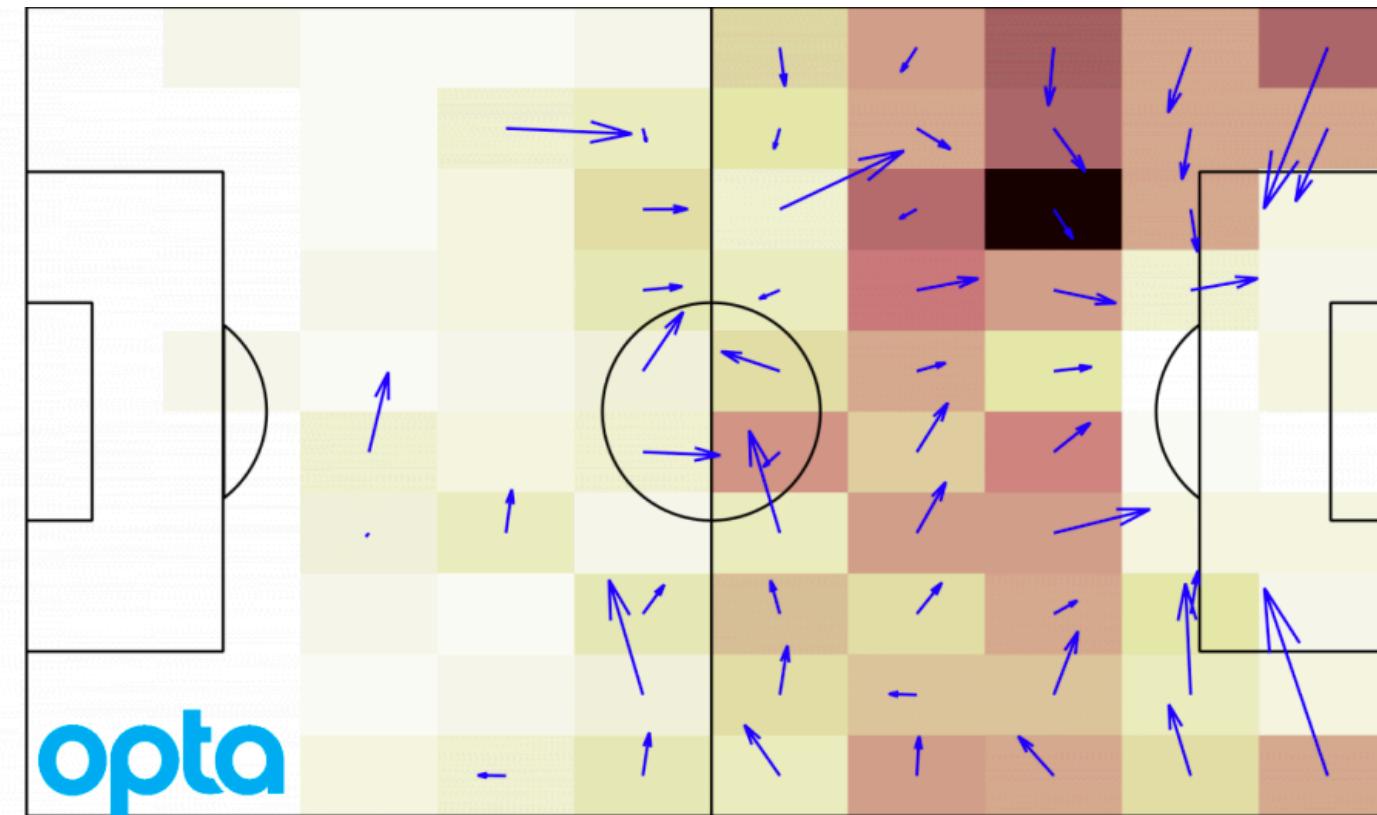


Touch heat maps are less useful.



Context is important

# Pass heat maps



Mesut Özil  
(2015)

<https://www.fourfourtwo.com/features/soccermatics-how-mesut-ozil-so-good-and-why-wenger-relies-ramsey>

# Heat maps

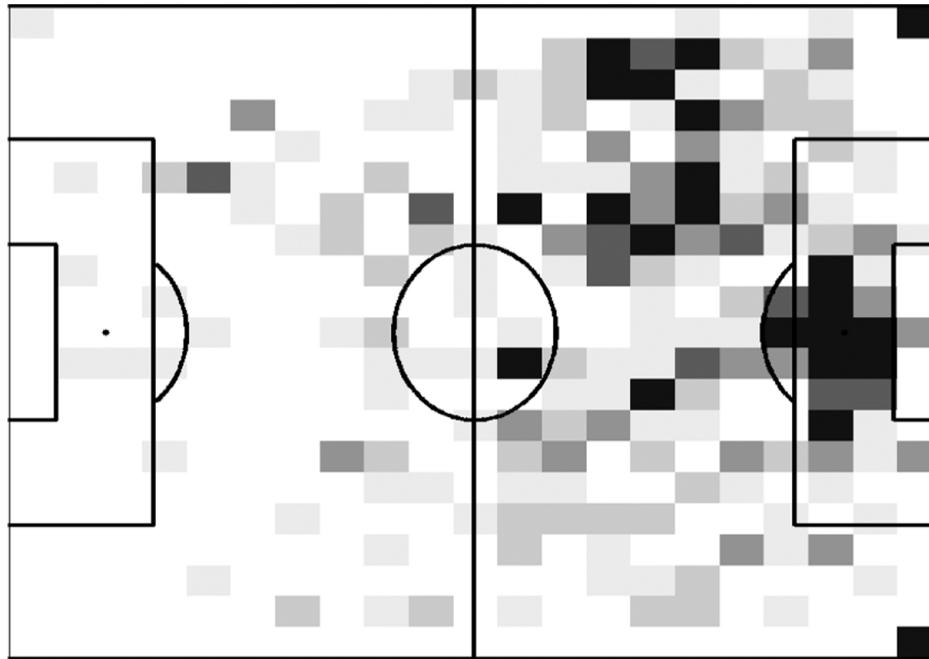
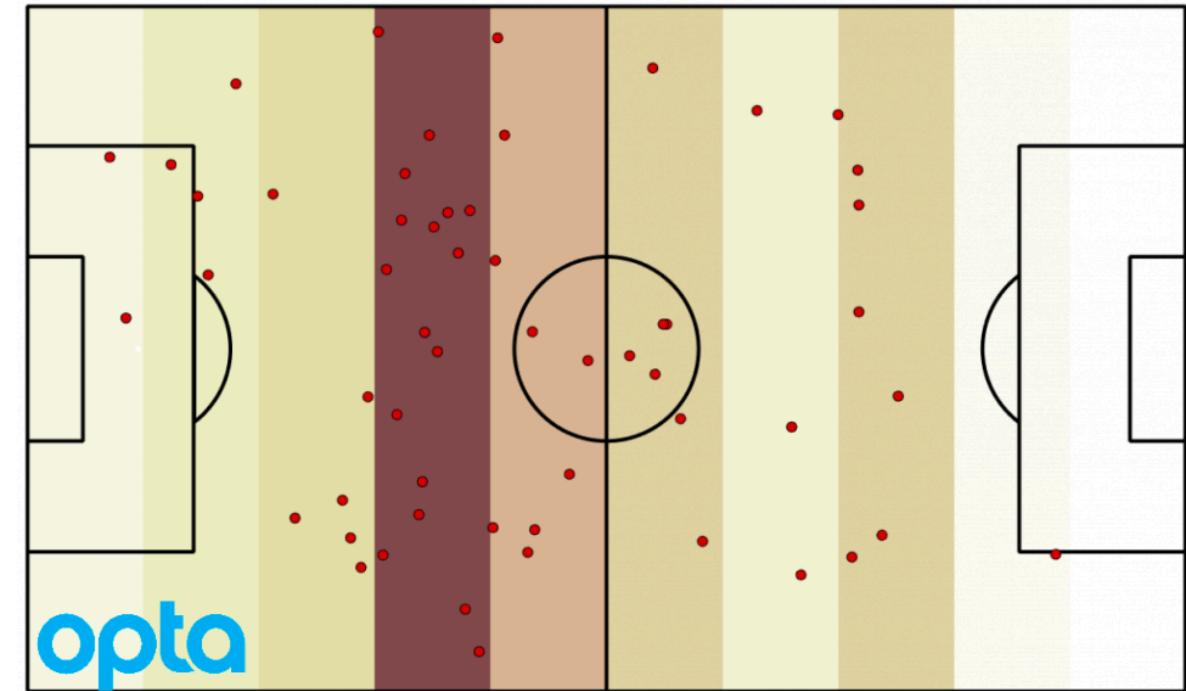
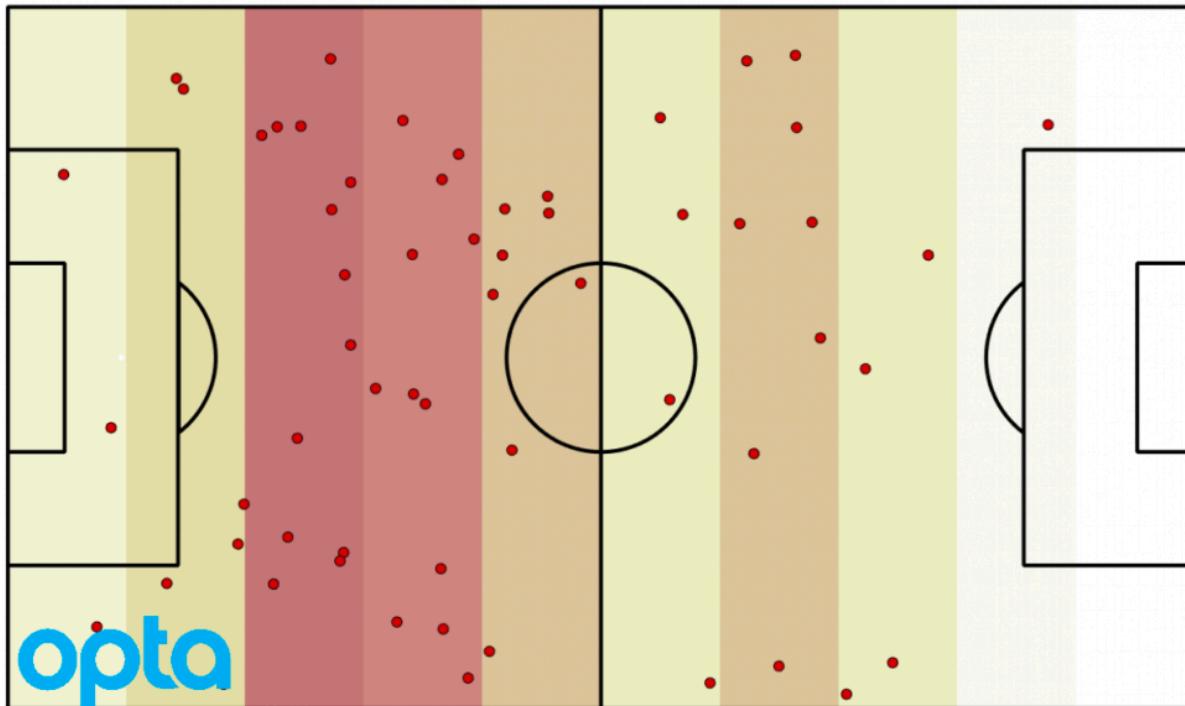


Figure 7.12 Real Madrid danger-zones during the Champions League season 2014/15. The shading is proportionally darker in areas where the ball was located during the 15 seconds leading up to a shot from the 20m by 20m area in front of the opposition goal. Data provided by Opta.

# Defensive actions



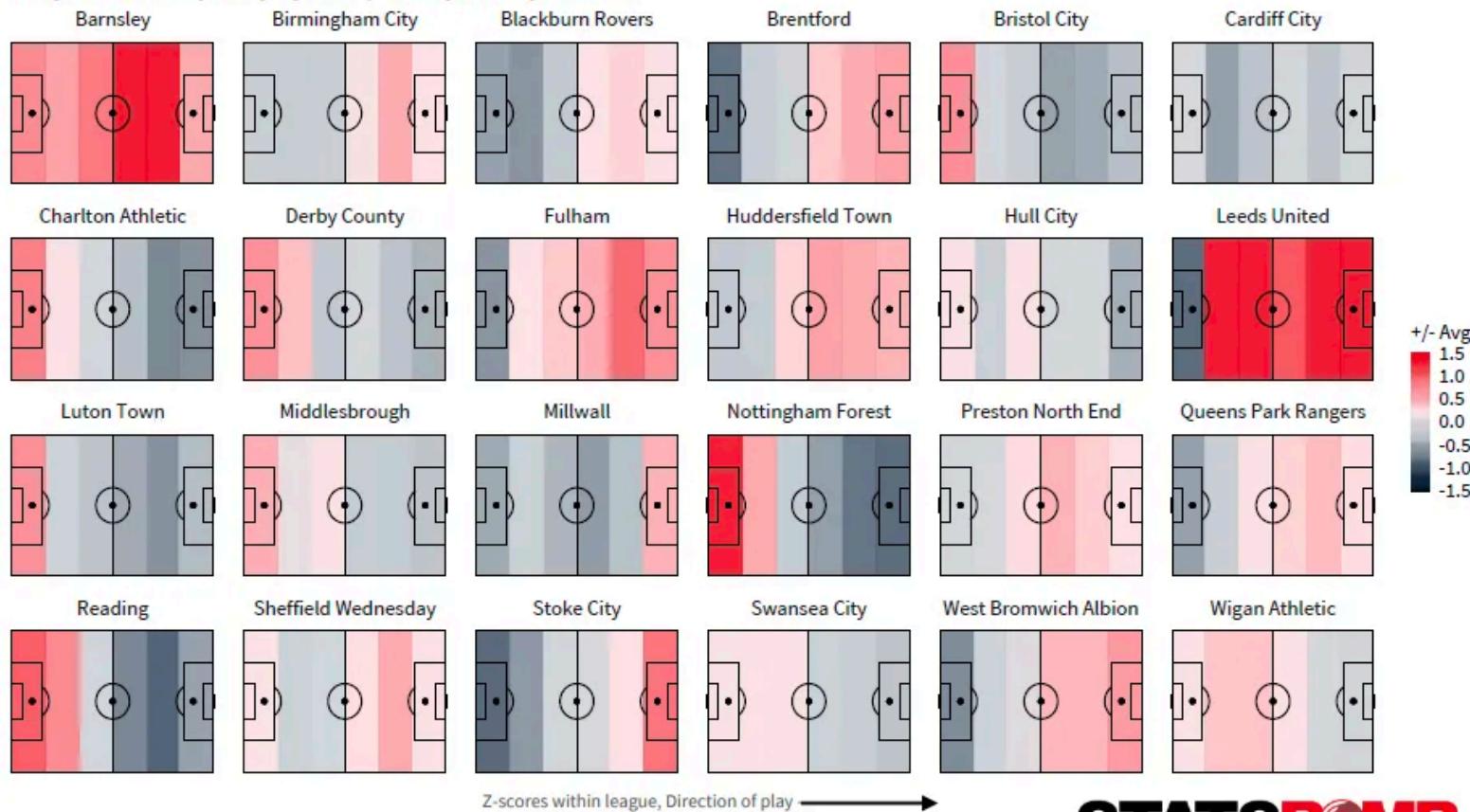
Revealed: Klopp's transition is more

important than the press <https://www.fourfourtwo.com/features/revealed-klopps-transition-more-important-press>

# Defensive actions

Proportion of Defensive Actions to Opposition Passes

Comparative Frequency by Zone, Championship 2019-20



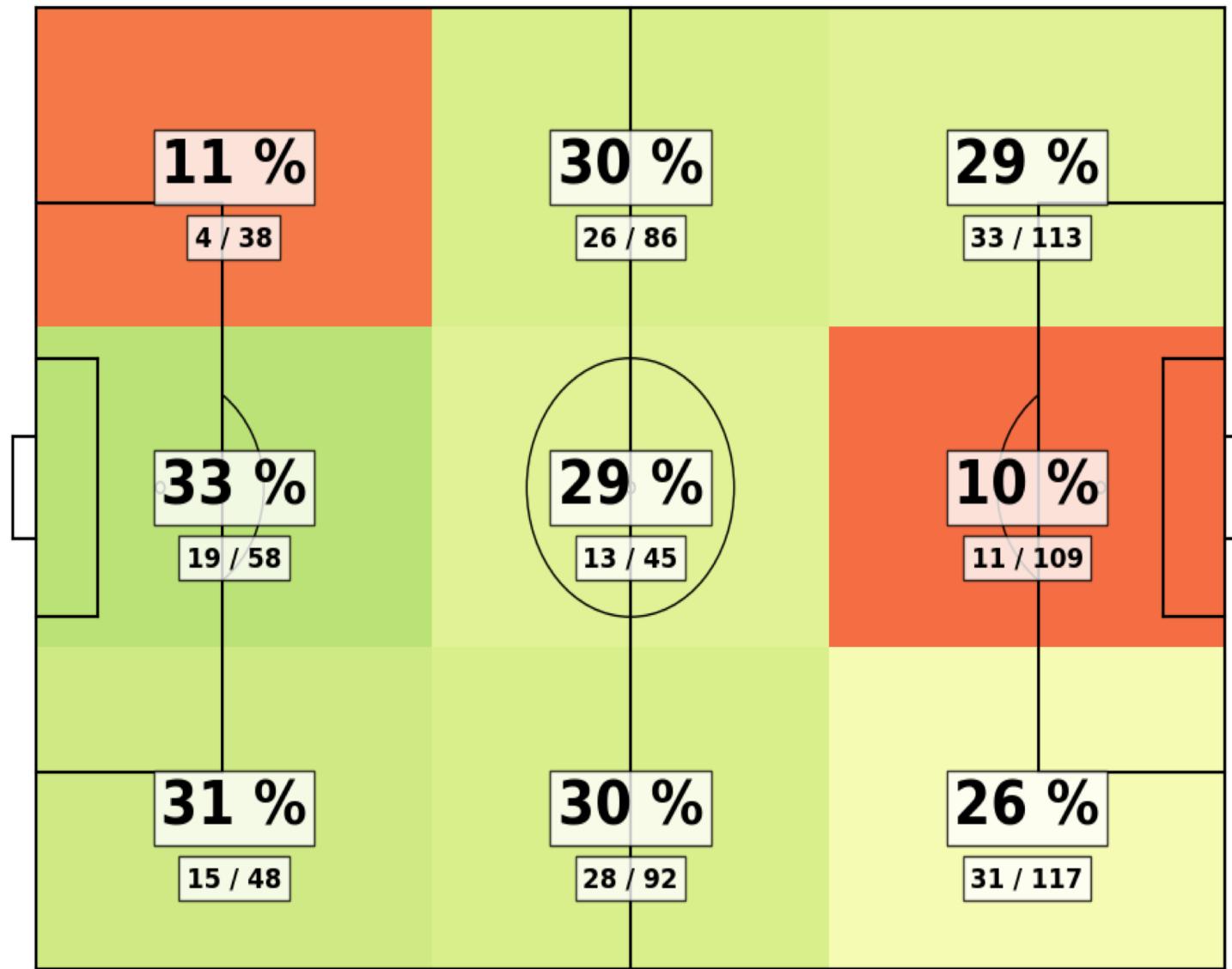
**STATSBOMB**

# Key Performance indices at clubs

- Entries final third/box.
- Shots danger zone.
- Number of passes leading to shot.
- Ball recoveries within 5 seconds.
- Passing tempo.
- Expected goals.

Single set of measures that are used over the entire club.

### Ball recoveries within 5 seconds, Hammarby IF



# Writing using statistics

- Tom Worville (e.g. <https://theathletic.co.uk/1956276/2020/07/29/bruno-fernandes-manchester-united-passing-risk-reward/>) requires Athletic subscription.
- James Yorke (e.g. <https://statsbomb.com/2019/08/anatomy-of-a-half/>).
- Nico Morales  
(<https://www.theringer.com/soccer/2018/7/9/17547678/2018-world-cup-semifinal-preview-john-stones-toby-alderweireld-samuel-umtiti>)

And many more....

# Summary

- Raw data is seldom enough.
- Standard visualisations are seldom enough (why we need Python!)
- What is the question? How do you answer it?
- Does the data support your hypothesis?
- Danger of self-confirmation, but risk of ignoring domain knowledge.
- Building measures that can then be used for future benchmarking (KPIs)
- Use both deductive (story) vs. inductive (data) thinking.