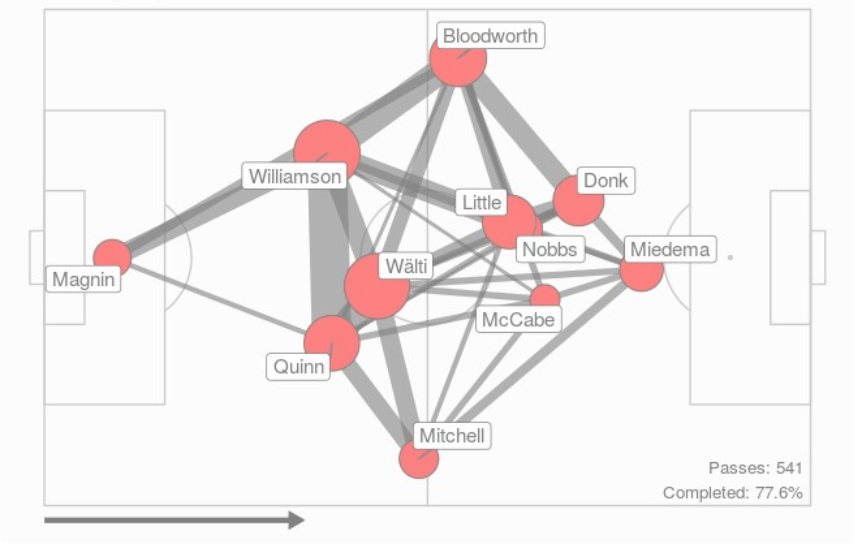
Football analytics with the **soccermatics** package

Joe Gallagher



jogall/soccermatics

@joedgallagher

LIVERPOOL FC

TIMES
DAY APRIL 3 2019

MOHAMED SALAH

Mohammed Salah: builds hospitals, feeds the poor, boosts Egypt's economy

Jonathan Northcroft, Football correspondent

December 17 2017, 12:01am, The Sunday Times

TIMES
AY APRIL 3 2019

MOHAMED SALAH

Mohammed Salah: builds hospitals, feeds the poor, boosts Egypt's economy

Jonathan Northcroft, Football correspondent

December 17 2017, 12:01am, The Sunday Times

England failure steps off plane and insults fans by showing off blinging house

EXCLUSIVE by RICHARD MORIARTY

FLOP Raheem Sterling faced fury last night for flashing a show of wealth in a video posted hours after flying home from England's Euro 2016 shame.

The Manchester City winger, 22, highlighted a crystal-encrusted bathroom sink during a tour of a mansion.

Full Story — Page Five

OBSCENE RAHEEM

TIMES
AY APRIL 3 2019

MOHAMED SALAH

Mohammed Salah: l
the poor, boosts Egy

Jonathan Northcroft, Football correspondent

December 17 2017, 12:01am, The Sunday Times

Mail Online

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money

Football | Premier League | Champions League | Rugby | Boxing | F1 | Tennis | MMA | Cricket | Golf | Racin

Alexis Sanchez sets up an Instagram page for his dogs... and it attracts almost 13,000 followers in a day as he reveals he has their names stitched on his BOOTS

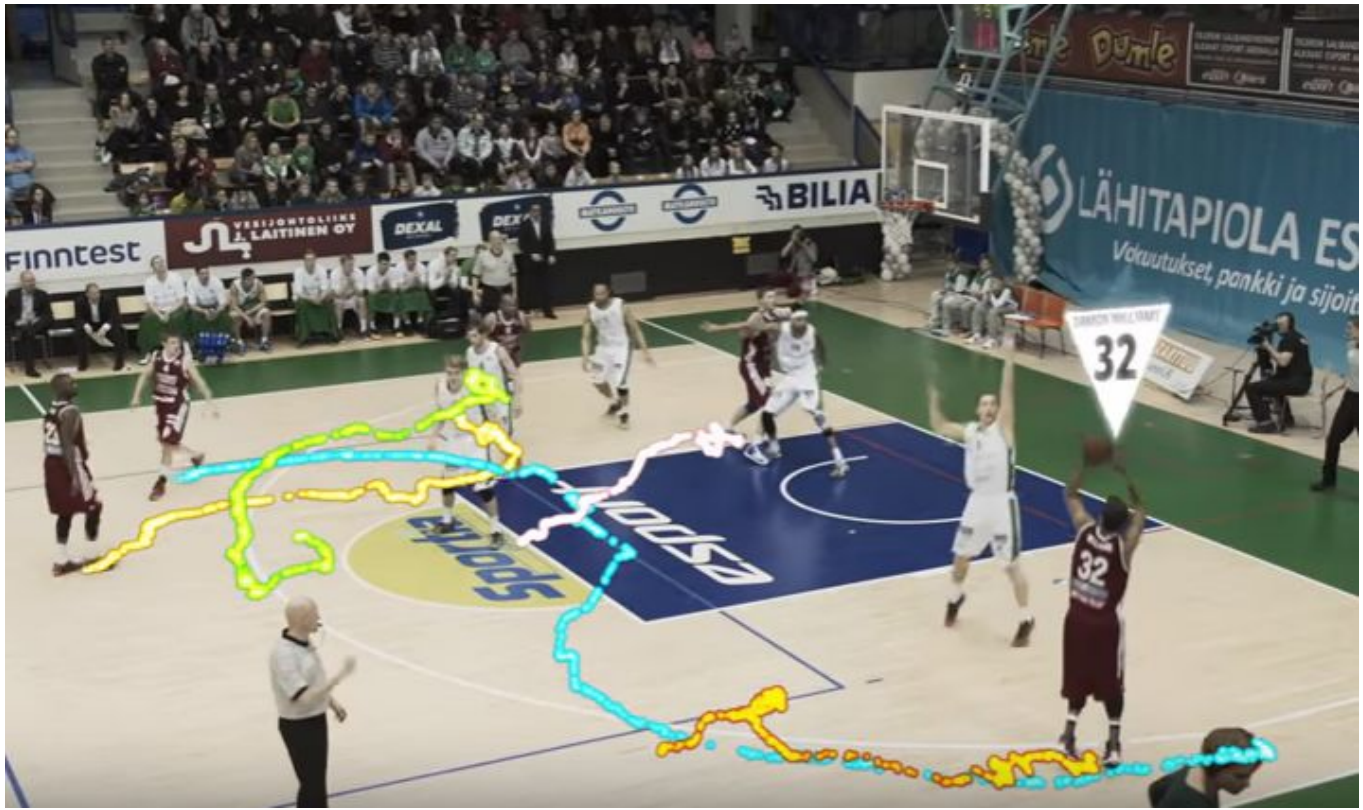
BSCENE
AHEEM

Analytics

Analytics

Equations
2x = −64
√(4.
T = 2
Schrödinger's Bat
Just Physics
BILL JAMES
MONEYBALL
F=MA
e^(iπ)+1=0
STATS
FOX

Analytics



Analytics

Line-ups
Match Stats
Live Text
Wolves
Man Utd
Possession
50%
50%
Shots
9
18
Shots on Target
2
4
Corners
3
5
Fouls
5
11

Analytics

Analytics

Challenges

Challenges

- Different x,y-coordinate systems
  - Opta: $0 < x < 100$, $0 < y < 100$
  - Statsbomb: $0 < x < 120$, $0 < y < 80$
  - TRACAB: $-5250 < x < 5250$, $-3400 < y < 3400$

Challenges

- Different x,y-coordinate systems
  - Opta: $0 < x < 100$, $0 < y < 100$
  - Statsbomb: $0 < x < 120$, $0 < y < 80$
  - TRACAB: $-5250 < x < 5250$, $-3400 < y < 3400$

- Different field names
  - e.g. ‘x’, ‘location.x’, ‘start.x’, etc...

Challenges

- Different x,y-coordinate systems
  - Opta: $0 < x < 100$, $0 < y < 100$
  - Statsbomb: $0 < x < 120$, $0 < y < 80$
  - TRACAB: $-5250 < x < 5250$, $-3400 < y < 3400$
- Different field names
  - e.g. ‘x’, ‘location.x’, ‘start.x’, etc...
- Different event logging
  - e.g. only StatsBomb track ‘Pressure’ events

Challenges

- Different x,y-coordinate systems
  - Opta: $0 < x < 100$, $0 < y < 100$
  - Statsbomb: $0 < x < 120$, $0 < y < 80$
  - TRACAB: $-5250 < x < 5250$, $-3400 < y < 3400$
- Different field names
  - e.g. ‘x’, ‘location.x’, ‘start.x’, etc...
- Different event logging
  - e.g. only StatsBomb track ‘Pressure’ events
- Different quantifications
  - e.g. what constitutes a key pass, or a big chance?

Challenges

- ACCESS TO DATA!

Challenges

- ACCESS TO DATA!

understat.com

AMERICAN SOCCER ANALYSIS

Challenges

- ACCESS TO DATA!

understat.com

Analytics

Challenges

- ACCESS TO DATA!

understat.com

jsonlite

XML

RSelenium

Free data!

Free data!

```r
# get free FA WSL data from StatsBomb API
fixtures <- StatsBombR::FreeMatches(37)
dat <- StatsBombR::getallinfo(fixtures[1,])
```

StatsBombR

Processing data

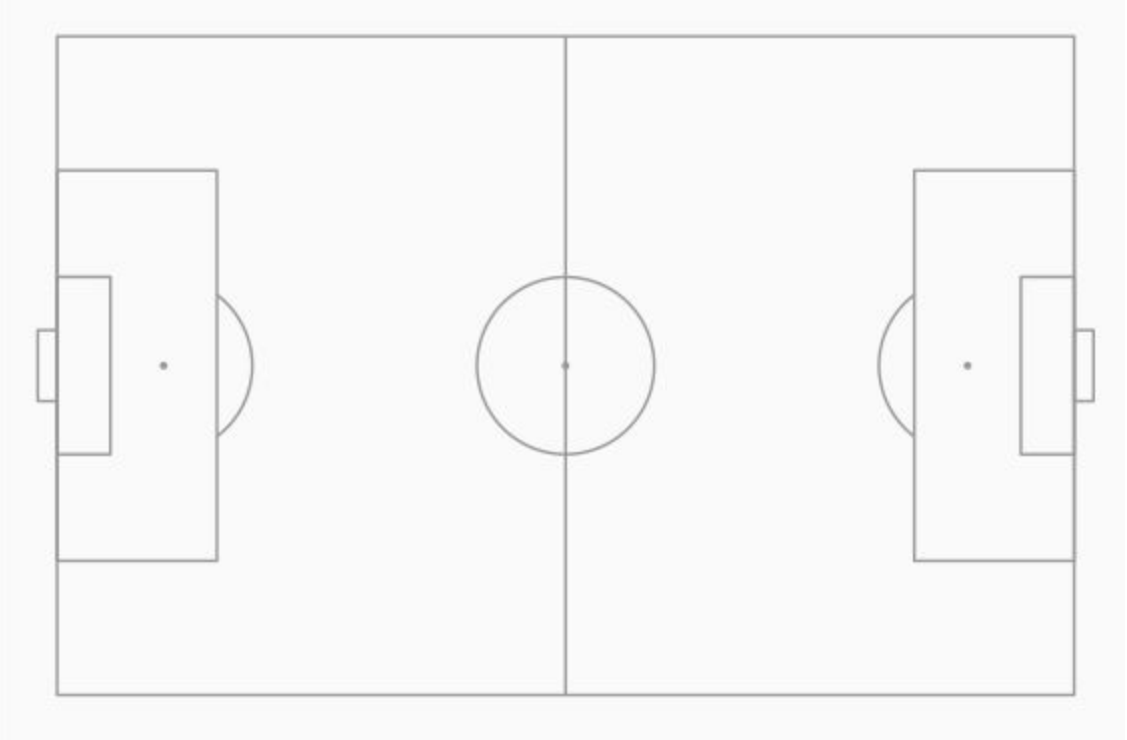
```
# convert x,y-coordinates for use with soccermatics
dat <- dat %>%
  soccerTransform(lengthPitch = 105, widthPitch = 68,
                  method = "statsbomb")
```

soccerTransform()

tidyverse

Pitch markings

```
soccerPitch(lengthPitch = 105, widthPitch = 68)
```



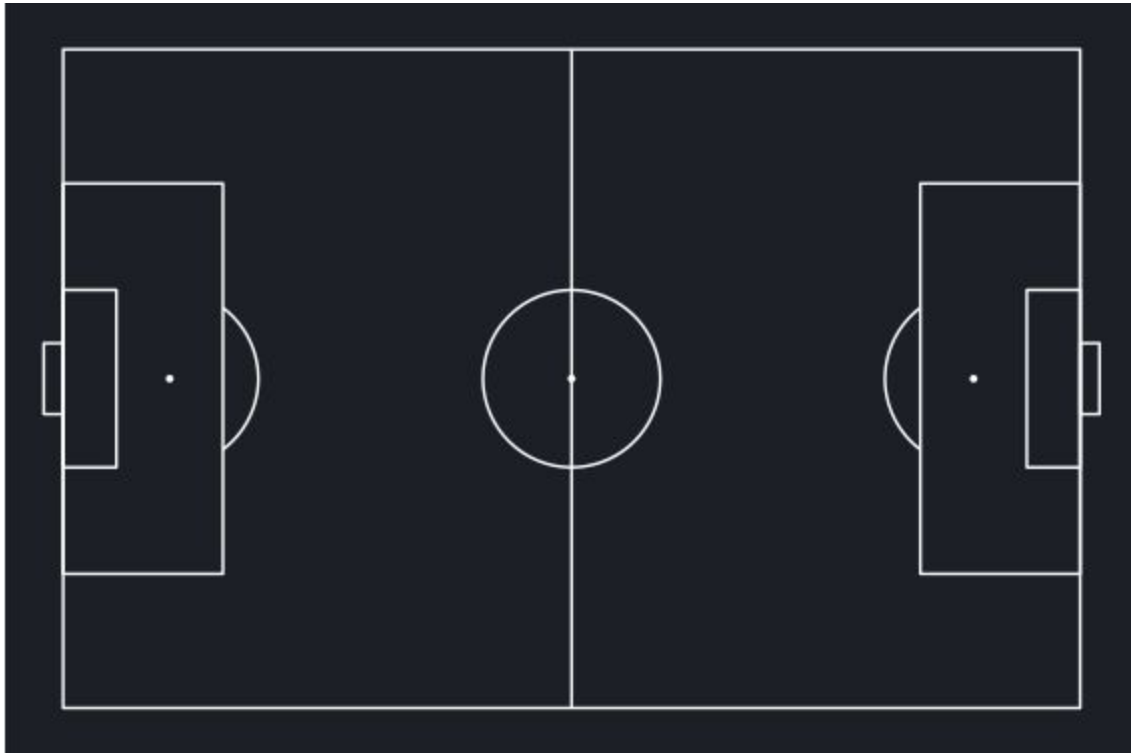
ggplot2

ggforce

soccermatics

Pitch markings

```
soccerPitch(theme = "dark")
```

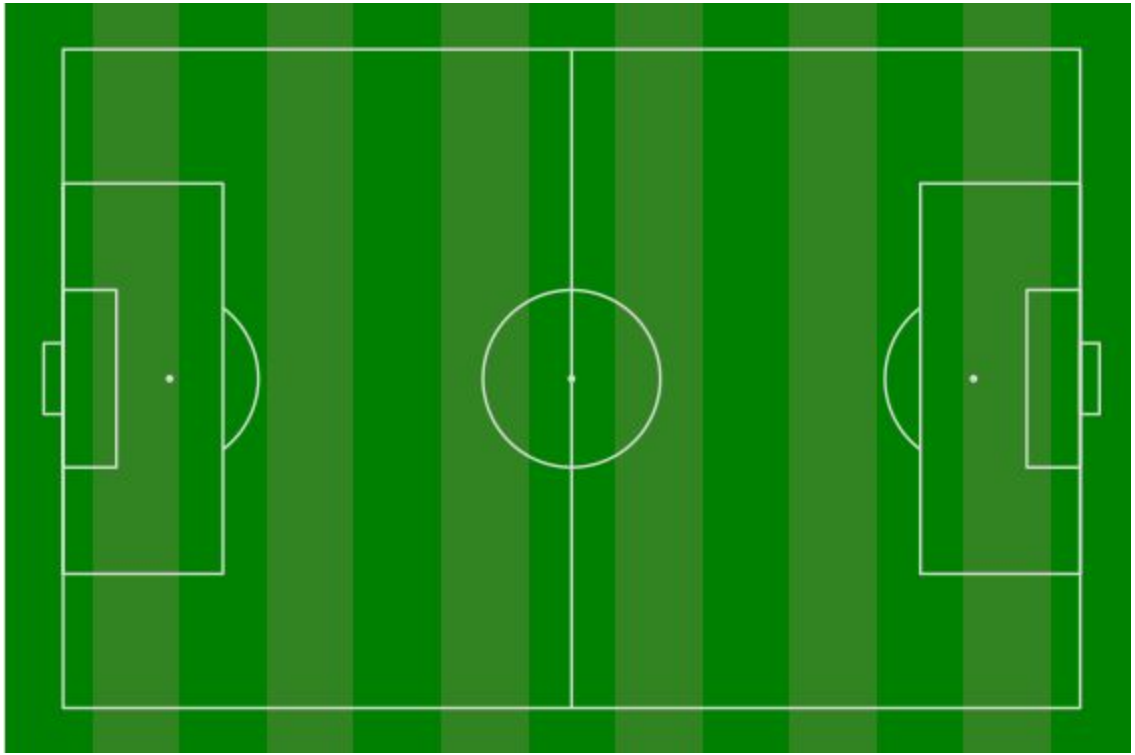


ggplot2

ggforce

Pitch markings

```
soccerPitch(theme = "grass")
```

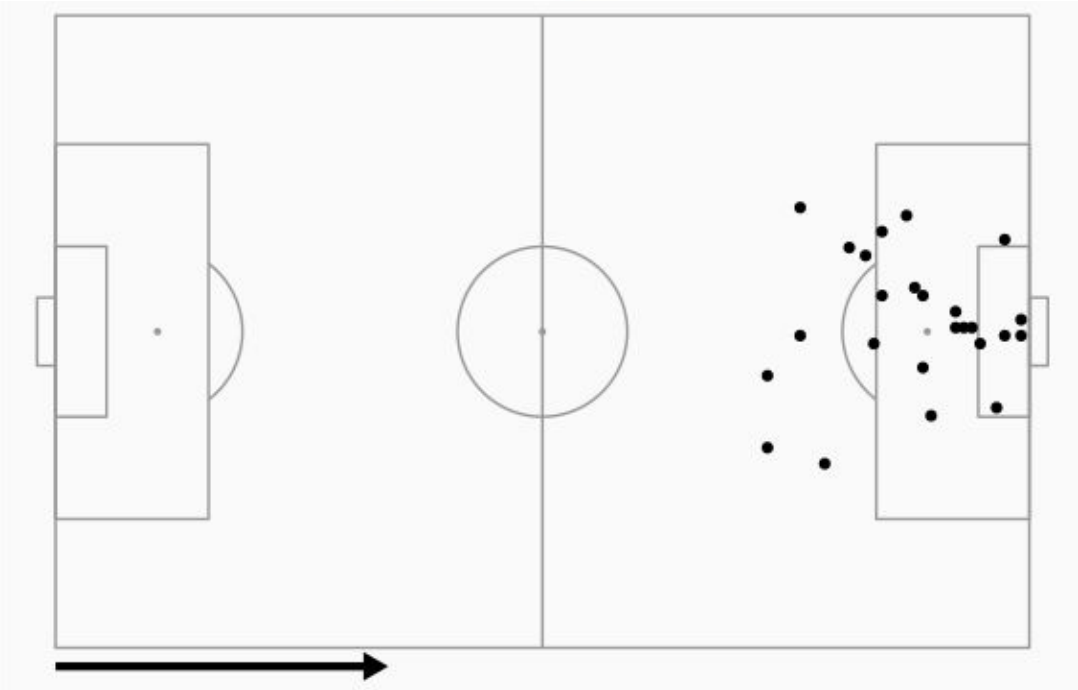


ggplot2

ggforce

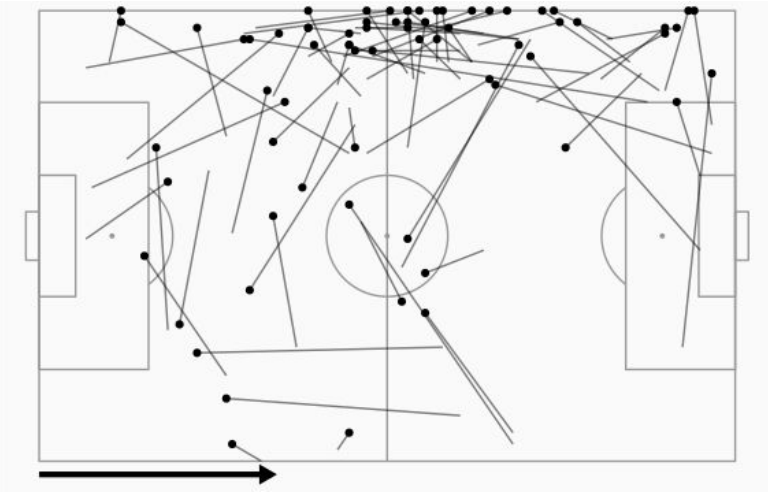
Simple visualisations

```
# plot shot locations
soccerPitch(arrow = "r") +
  geom_point(data = dat %>%
               filter(type.name == "Shot" & team.name == "Arsenal WFC"),
             aes(location.x, location.y))
```



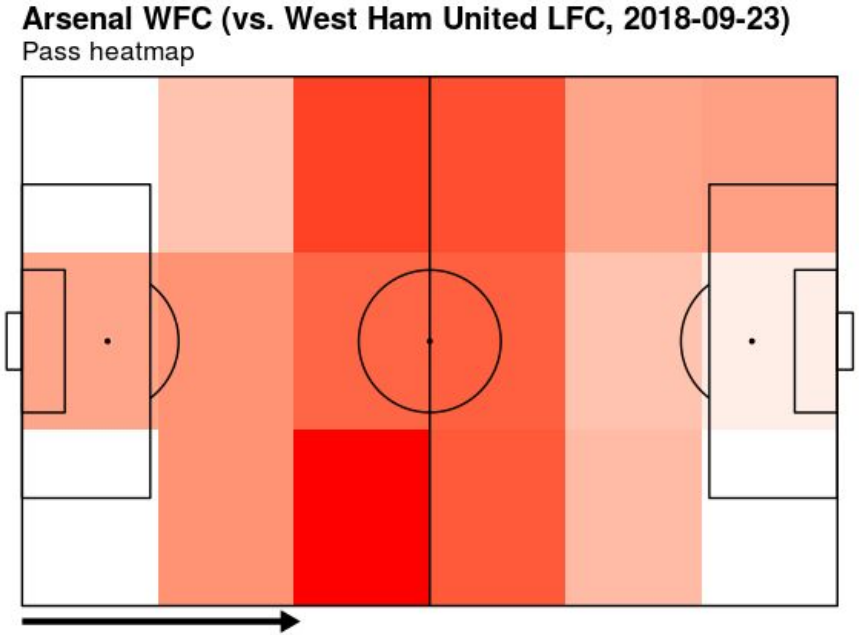
Simple visualisations

```
# plot pass start and end locations
dat %>%
  filter(type.name == "Pass" & player.name == "Dominique Johanna Anna Bloodworth") %>%
  {
    soccerPitch(arrow = "r") +
    geom_point(data = ., aes(x = location.x, y = location.y)) +
    geom_segment(data = ., aes(x = location.x, y = location.y,
                               xend = pass.end_location.x, yend = pass.end_location.y),
                 alpha = 0.5)
  }
```



Zonal heatmap

```
dat %>%
  filter(type.name == "Pass" & team.name == "Arsenal WFC") %>%
  soccerHeatmap(x = "location.x", y = "location.y", xBins = 6, yBins = 3,
                arrow = "r",
                title = "Arsenal WFC (vs. West Ham United LFC, 2018-09-23)",
                subtitle = "Pass heatmap")
```

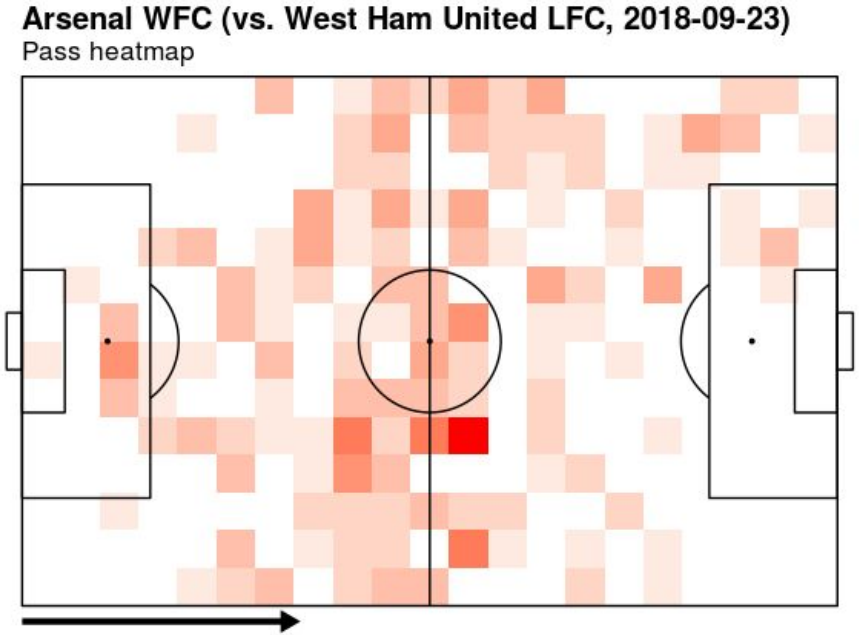


soccerHeatmap()

ggplot2::geom_bin2d()

Zonal heatmap

```
# ~ 5x5m bins
dat %>%
  filter(type.name == "Pass" & team.name == "Arsenal WFC") %>%
  soccerHeatmap(x = "location.x", y = "location.y", xBins = 21, yBins = 14,
                arrow = "r",
                title = "Arsenal WFC (vs. West Ham United LFC, 2018-09-23)",
                subtitle = "Pass heatmap")
```

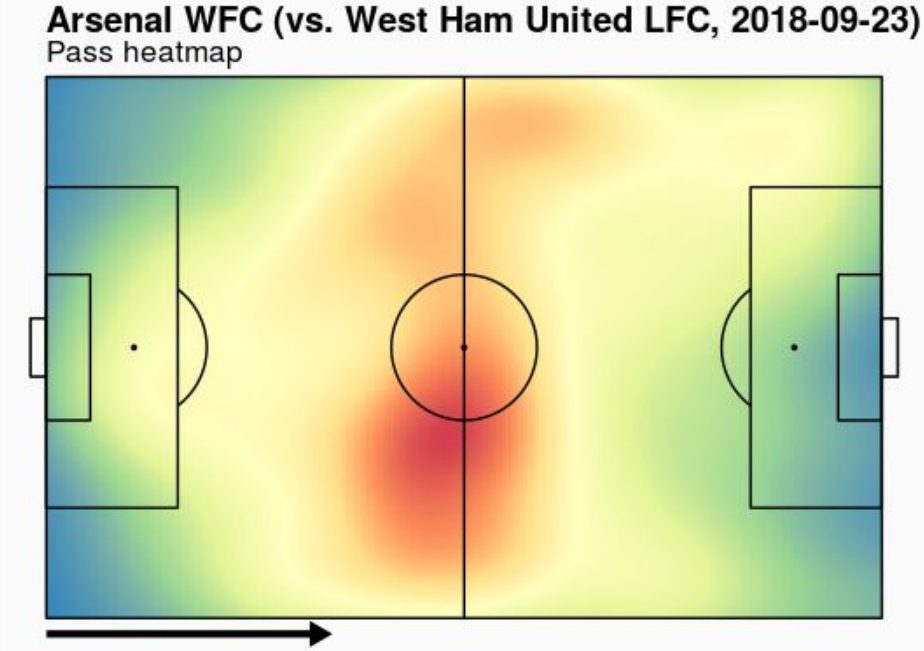


soccerHeatmap()

ggplot2::geom_bin2d()

KDE heatmap

```
dat %>%
  filter(type.name == "Pass" & team.name == "Arsenal WFC") %>%
  soccerHeatmap(x = "location.x", y = "location.y", kde = T,
                arrow = "r",
                title = "Arsenal WFC (vs. West Ham United LFC, 2018-09-23)",
                subtitle = "Pass heatmap")
```

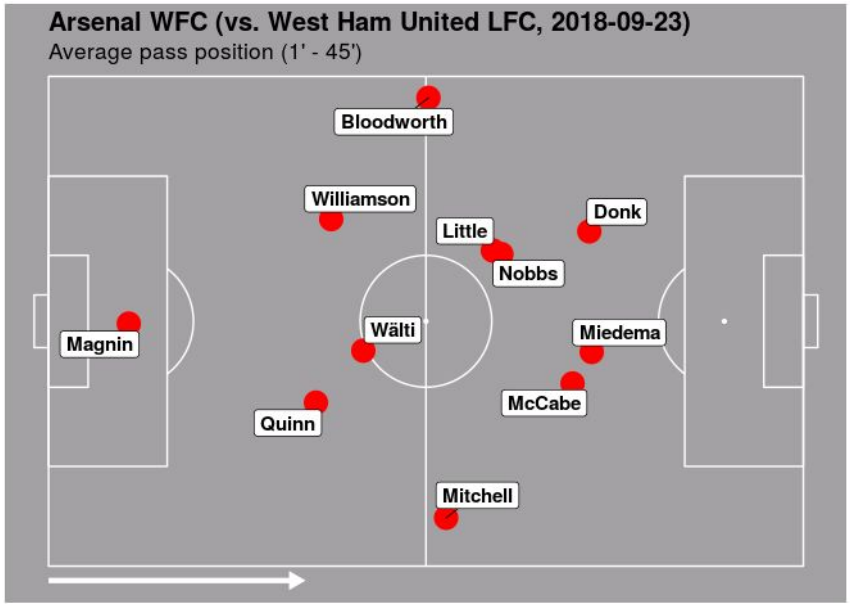


soccerHeatmap()

MASS::kde2d()

Average position

```
# average pass position for one team w/ player name labelled as text
dat %>%
  filter(type.name == "Pass" & team.name == "Arsenal WFC" & period == 1) %>%
  soccerPositionMap(source = "statsbomb",
                    fill1 = "red", arrow = "r", theme = "grey", labelCol = "white",
                    title = "Arsenal WFC (vs. West Ham United LFC, 2018-09-23)",
                    subtitle = "Average pass position (1' - 45')")
```



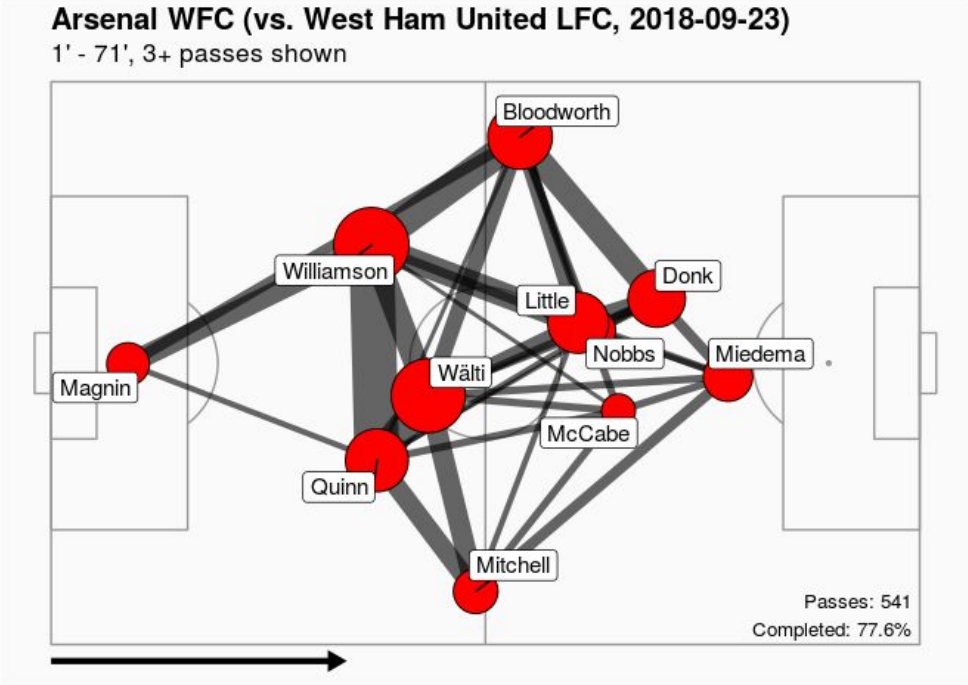
soccerPositionMap()

ggrepel

soccerShortenName()

Passmaps

```
dat %>%
  filter(team.name == "Arsenal WFC") %>%
  soccerPassmap(fill = "red", minPass = 3, arrow = "r",
                maxEdgeSize = 30,
                title = "Arsenal WFC (vs. West Ham United LFC, 2018-09-23)")
```



soccerPassmap()

Expected goals (xG)

soccermatics

Expected goals (xG)

Expected goals (xG)

Penalty = 76% chance of being scored

xG = 0.76

soccermatics

Expected goals (xG)

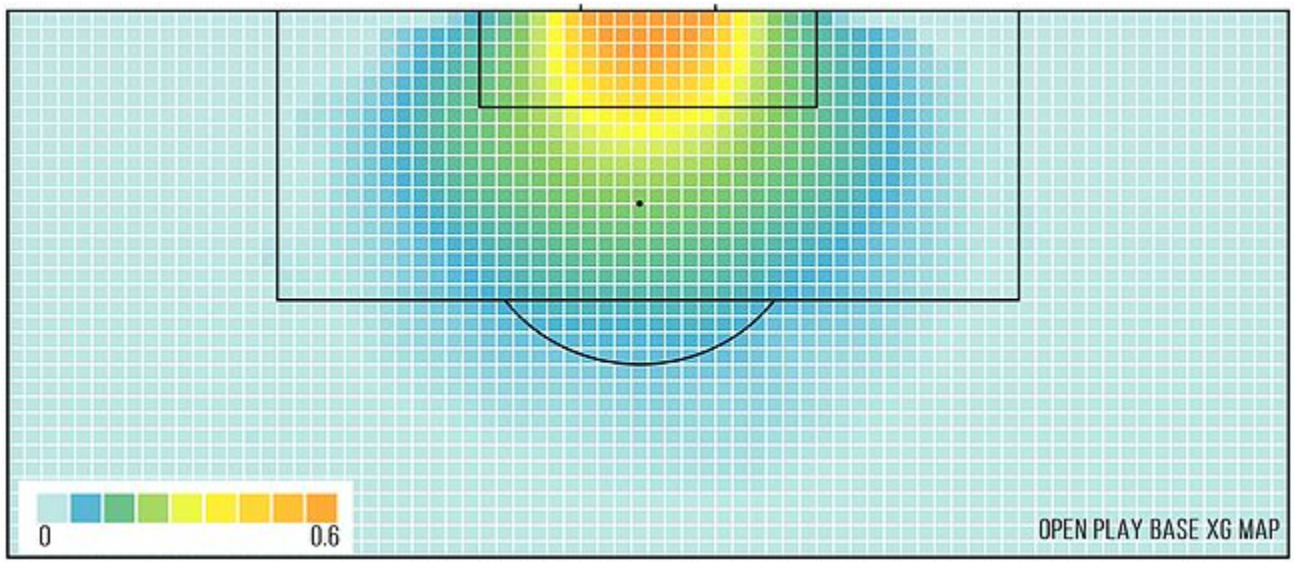
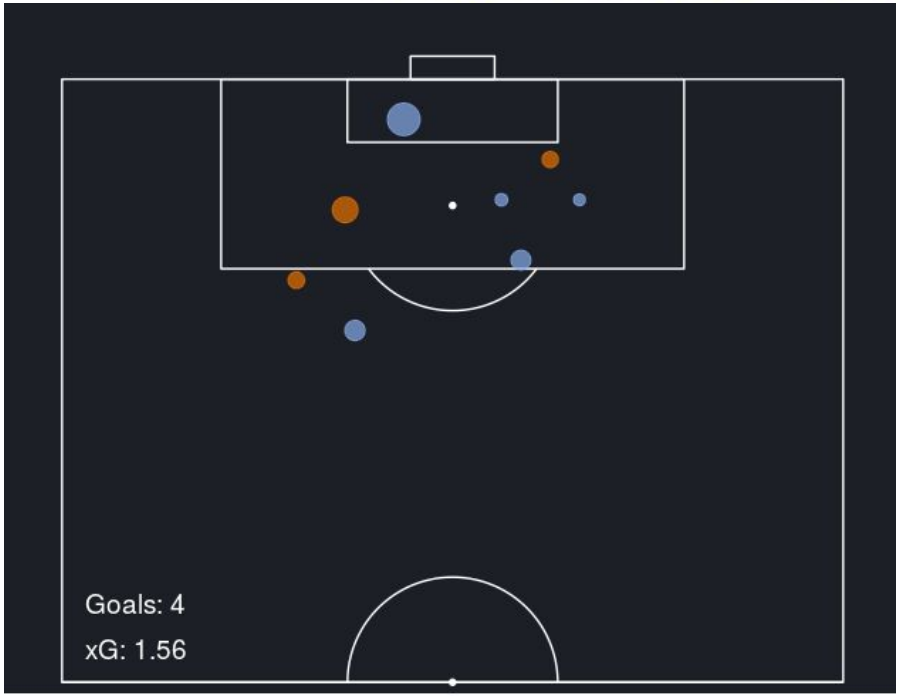


Image from:
https://s3-eu-west-1.amazonaws.com/scisports-production/2a8d8a_b0010c24664a4bd08a88c20f71256de3.png

Expected goals (xG)

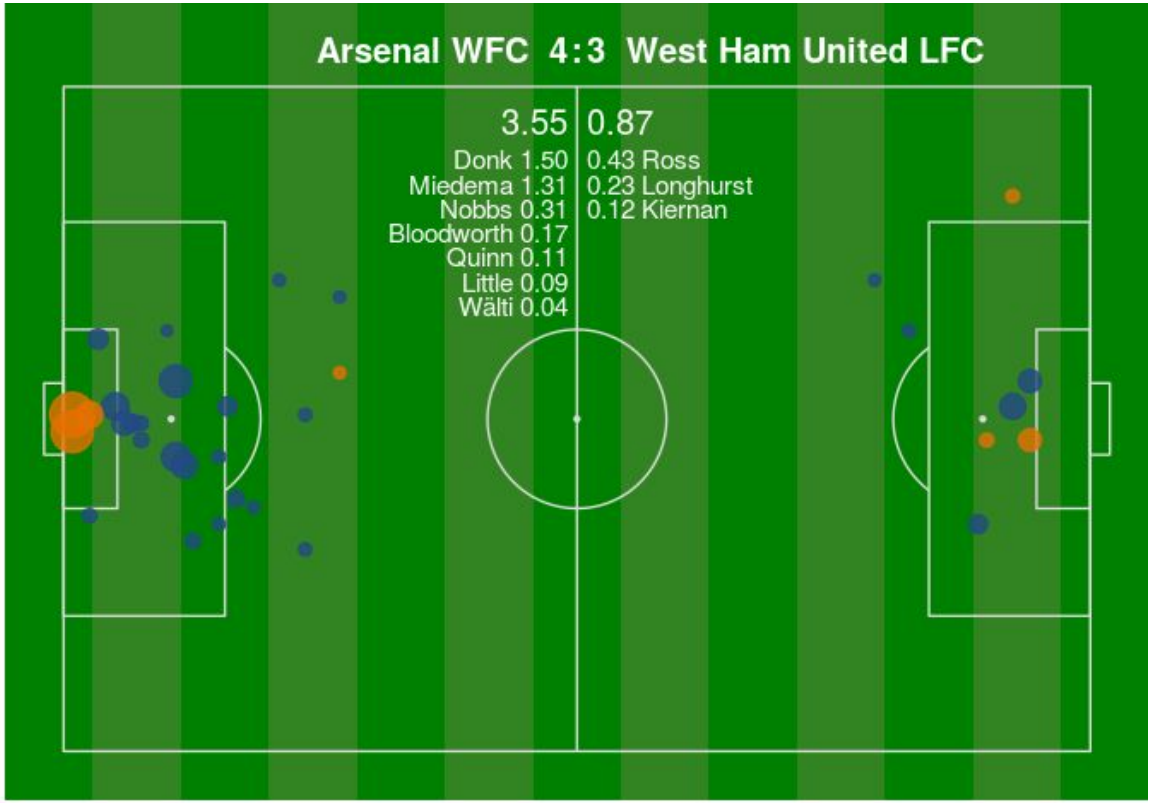
```
dat %>%
  filter(team.name == "Arsenal WFC") %>%
  soccerShotmap(theme = "dark",
                title = "Arsenal WFC",
                subtitle = "vs. West Ham United LFC (2018-09-23)")
```



soccerShotmap()

Expected goals (xG)

```
dat %>%
  soccerShotmap(theme = "grass", n_players = 10, adj = T)
```

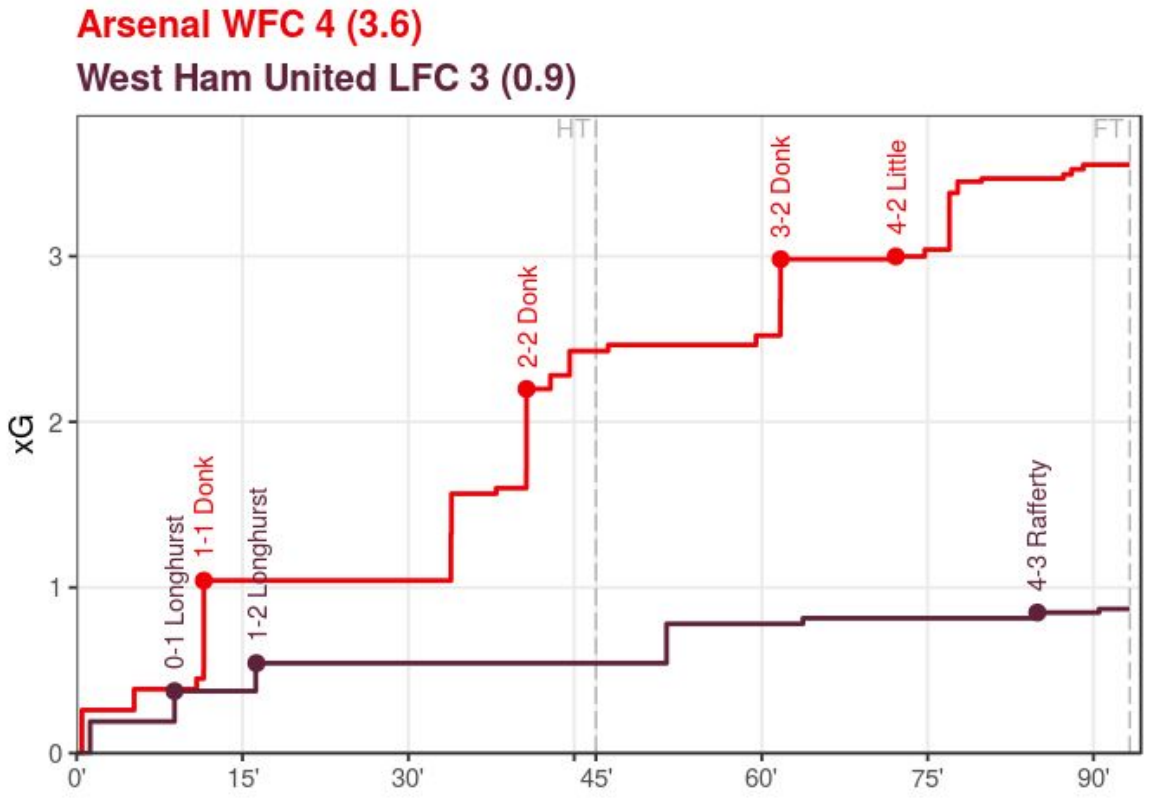


soccerShotmap()

soccermatics

Expected goals (xG)

```
dat %>%
  soccerxGTimeline(homeCol = "#EF0107", awayCol = "#60223B")
```



soccerxGTimeline()

Applications

@WomensFootyStat

Applications

@WomensFootyStat

WomensFootyStat @WomensFootyStat · Mar 18

Reading WFC 3 : 4 Manchester City WFC

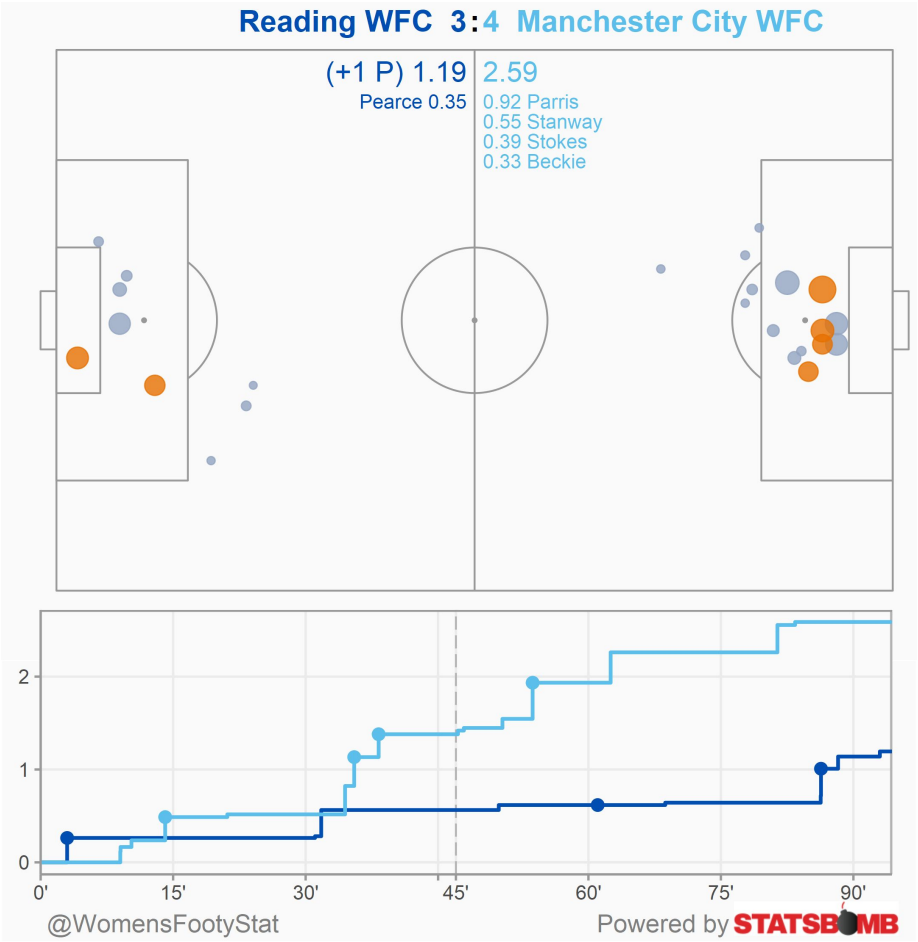
1.2 : 2.6 NPxG

xG: Parris (MCI)
Progressions: Stokes (MCI)
Defence: Houghton (MCI)

#xg #shotmap #passmap
#FAWSL #MCFC #RWFC
@ManCityWomen @ReadingFCWomen
@lilkeets @DemiStokes @stephhoughton2

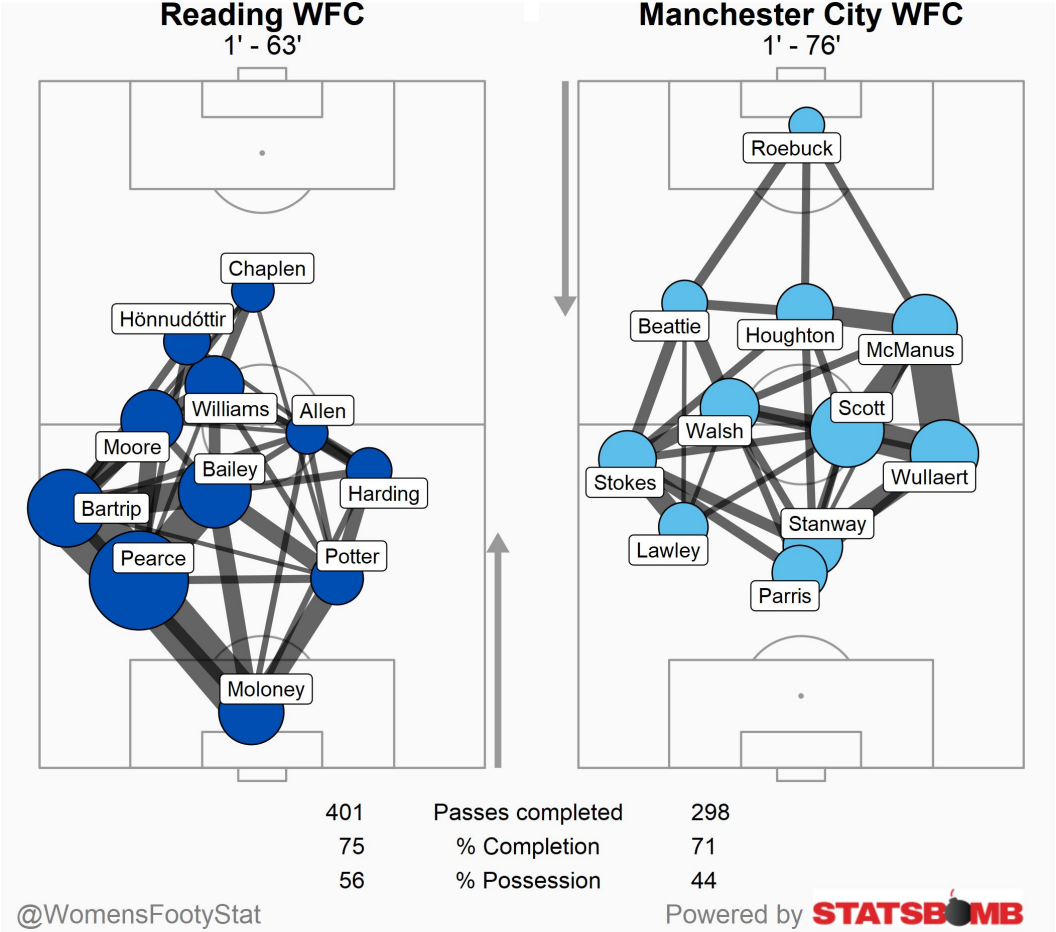
Applications

@WomensFootyStat



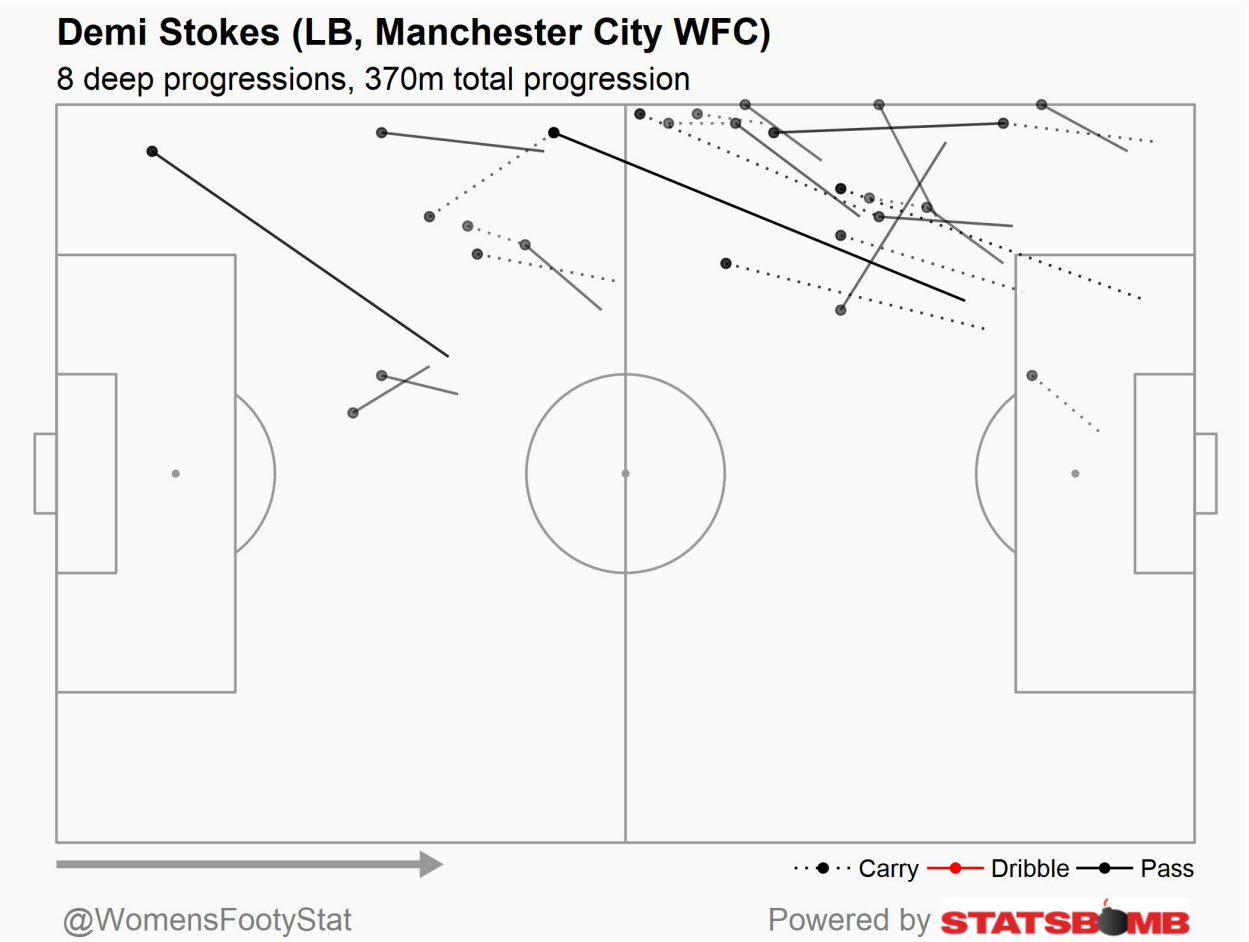
Applications

@WomensFootyStat

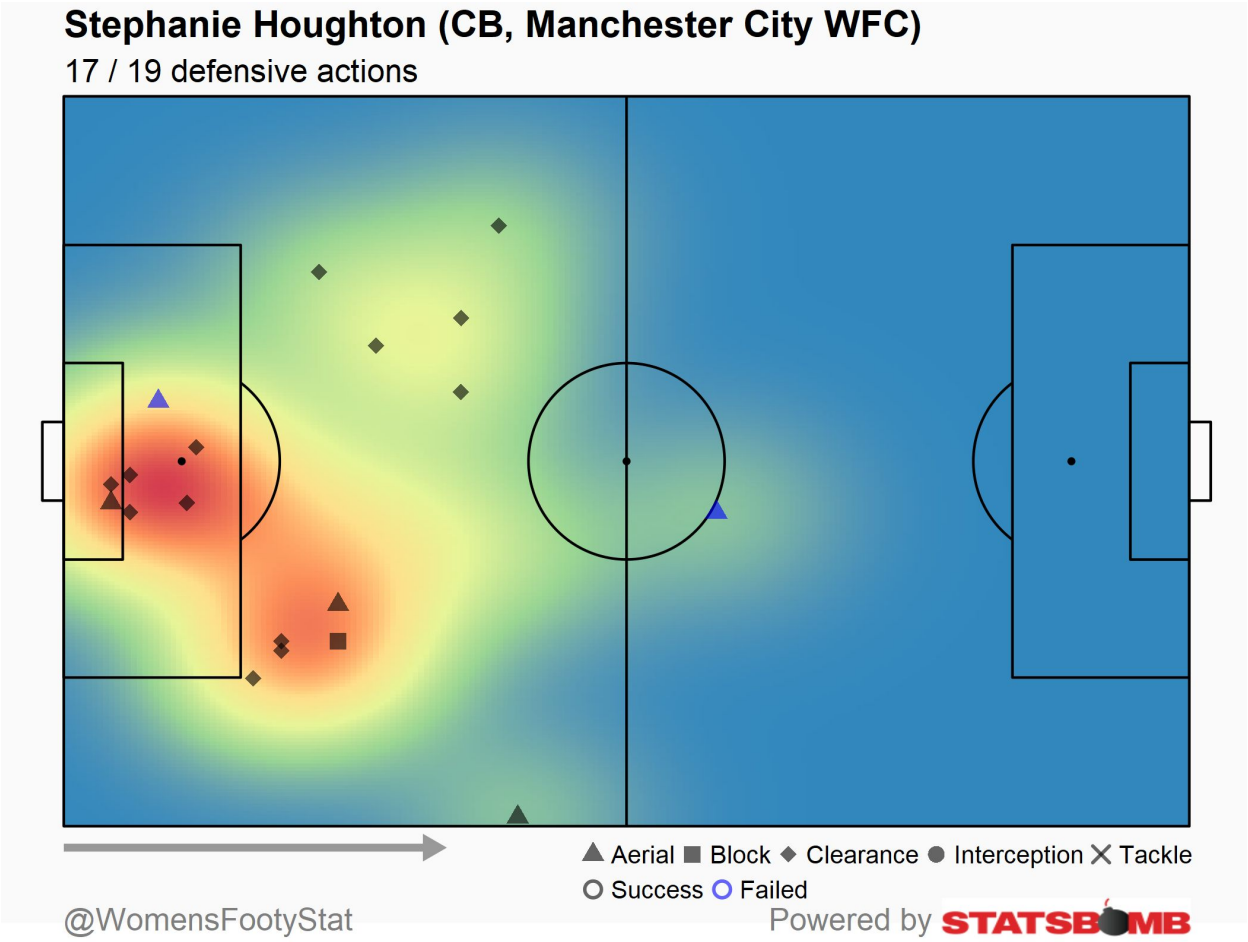


Applications

@WomensFootyStat



Applications



Applications

@WomensFootyStat

FAWSL: Tackles and interceptions

Possession-adjusted tackles and interceptions, per 30 minutes of opposition possession (near the average team possession per game)

(Minimum 800 minutes played; data as of 10th Feb 2019.)

#	Player	Team	Pos	Opp. Mins	# TI	PadjTI / 30
1	Angharad James	Everton LFC	CM	597	47	**2.36**
2	Leigh-Anne Robe	Liverpool WFC	LB	722	48	**1.99**
3	Kayleigh Green	Brighton & Hove Albion WFC	CM	573	37	**1.94**
4	Lia Wälti	Arsenal WFC	DM	651	42	**1.94**
5	Hayley Ladd	Birmingham City WFC	DM	682	43	**1.89**
6	Chloe Peplow	Brighton & Hove Albion WFC	CM	479	30	**1.88**
7	Jill Scott	Manchester City WFC	CM	736	46	**1.88**
8	Poppy Pattinson	Bristol City WFC	LB	594	37	**1.87**
9	Satara Murray	Liverpool WFC	RB	771	46	**1.79**
10	Danielle Buet	Brighton & Hove Albion WFC	DM	524	31	**1.77**

Applications

@WomensFootyStat

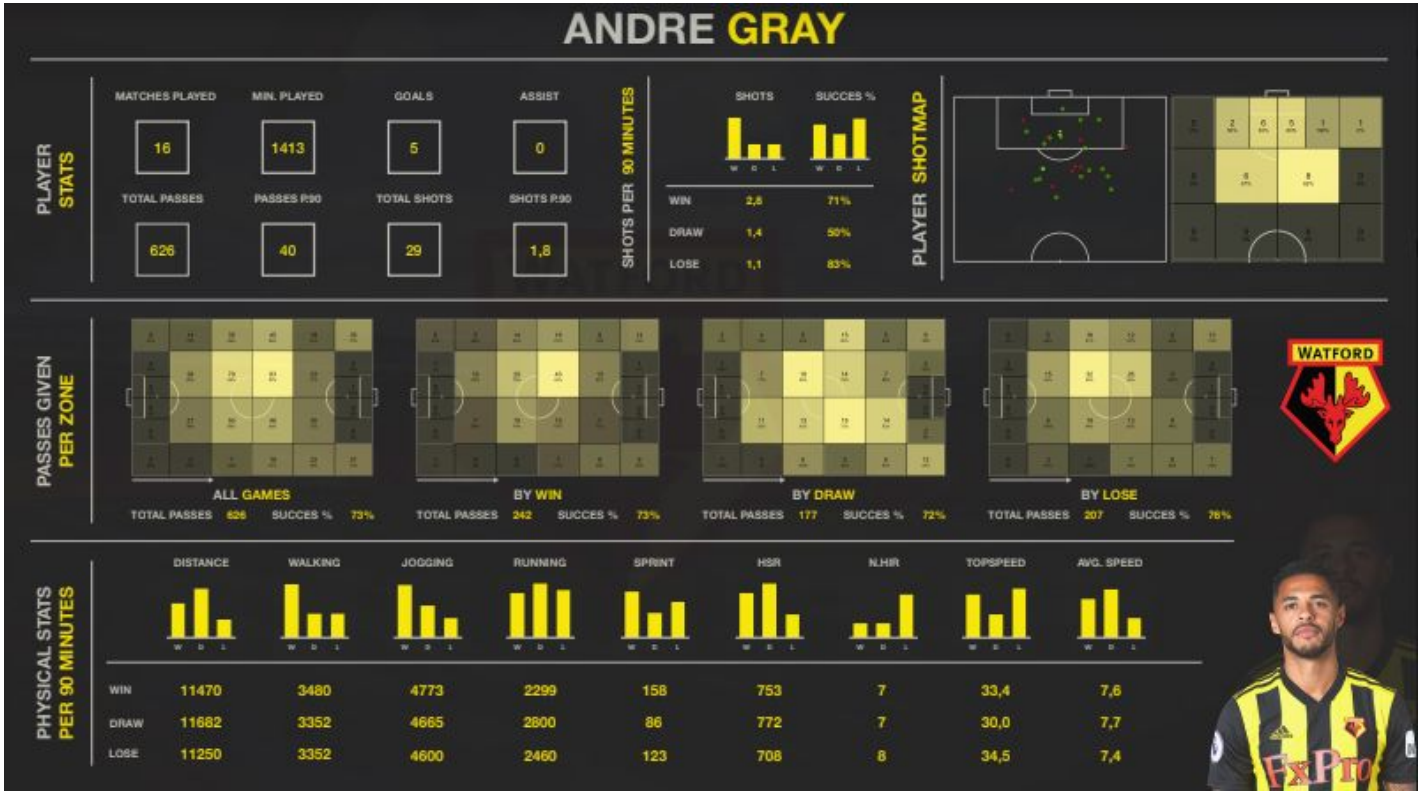
Automation!

gmailR

twittR

Applications

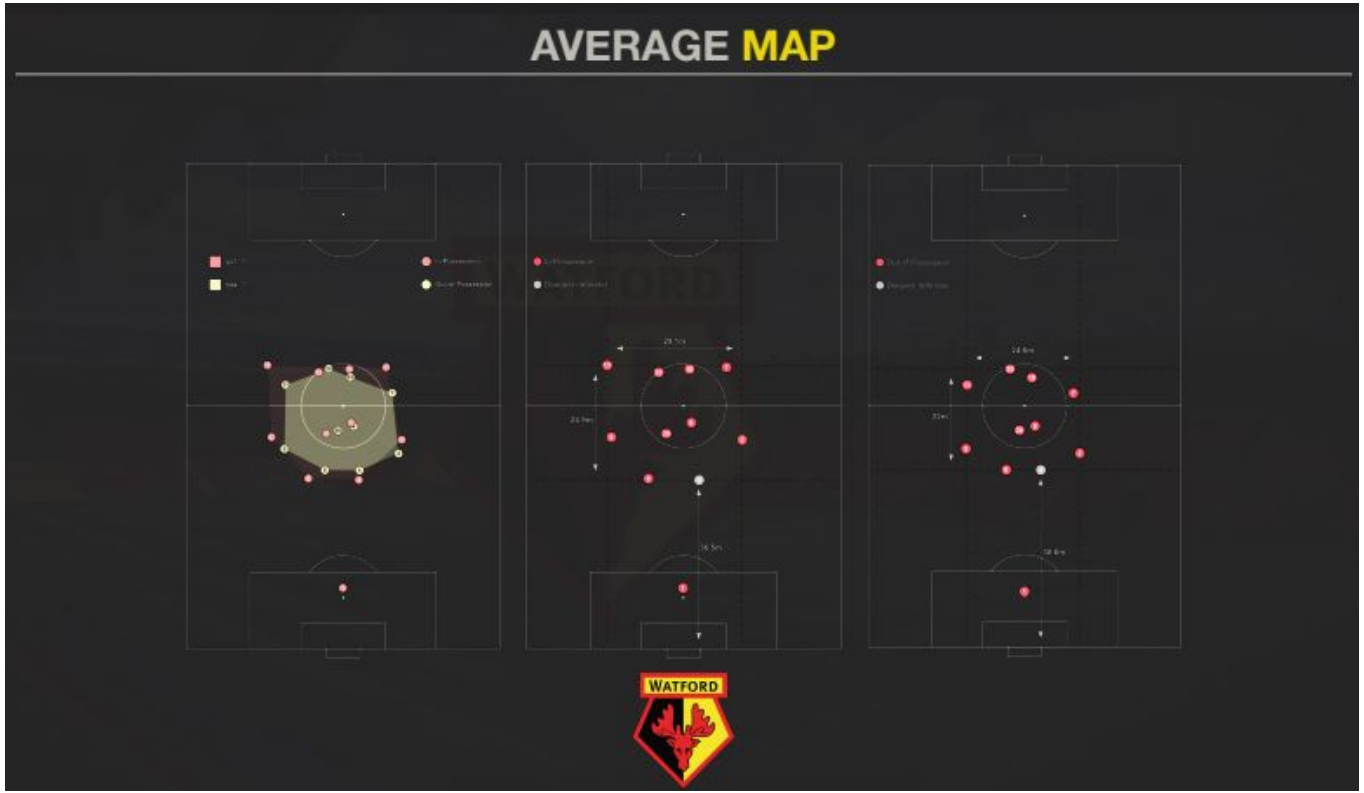
Opposition analysis



@JoeyHarings

Applications

Opposition analysis



@JoeyHarings

Future directions

Future directions

To do list #8

Open **JoGall** opened this issue on 1 Feb 2018 · 1 comment

JoGall commented on 1 Feb 2018 • edited ▾ Owner

A rough list of functions and utilities I'd like to add to the package. Recommendations welcomed!

🚧 = under construction
💡 = still just a pipedream

Visualisations

💡 Post-match odds based on simulating shots
🚧 Voronoi diagrams / Delaunay triangulation / convex hull
🚧 Wrapper functions to animate metrics frame by frame
💡 Defensive coverage plots

Advanced analytics

🚧 Wrapper functions for common metrics (xG, xA, forward progressions, PossAdjT&A, etc...)
💡 Event sequence plots and possession sequence metrics (possession spell and final product, i.e. does it end in final third, end in shot, end in goal)
💡 Directness, synchronisation of movement direction, space utilisation)
🚧 Space control / dominance (e.g. ratio of Voronoi areas / convex hull areas)
💡 Goalscoring threat (xG in each frame based on view to goal based on distance / angle of ball from

Future directions

Cleaner code!

I don't want your monolithic ggplot function

```
some_object_from_my_package %>%
  big_function_with_a_bunch_of_parameters_that_returns_a_ggplot(...)
```

@mjskay

Future directions

Cleaner code!

I don't want your monolithic ggplot function

```
some_object_from_my_package %>%
  big_function_with_a_bunch_of_parameters_that_returns_a_ggplot(...)
```

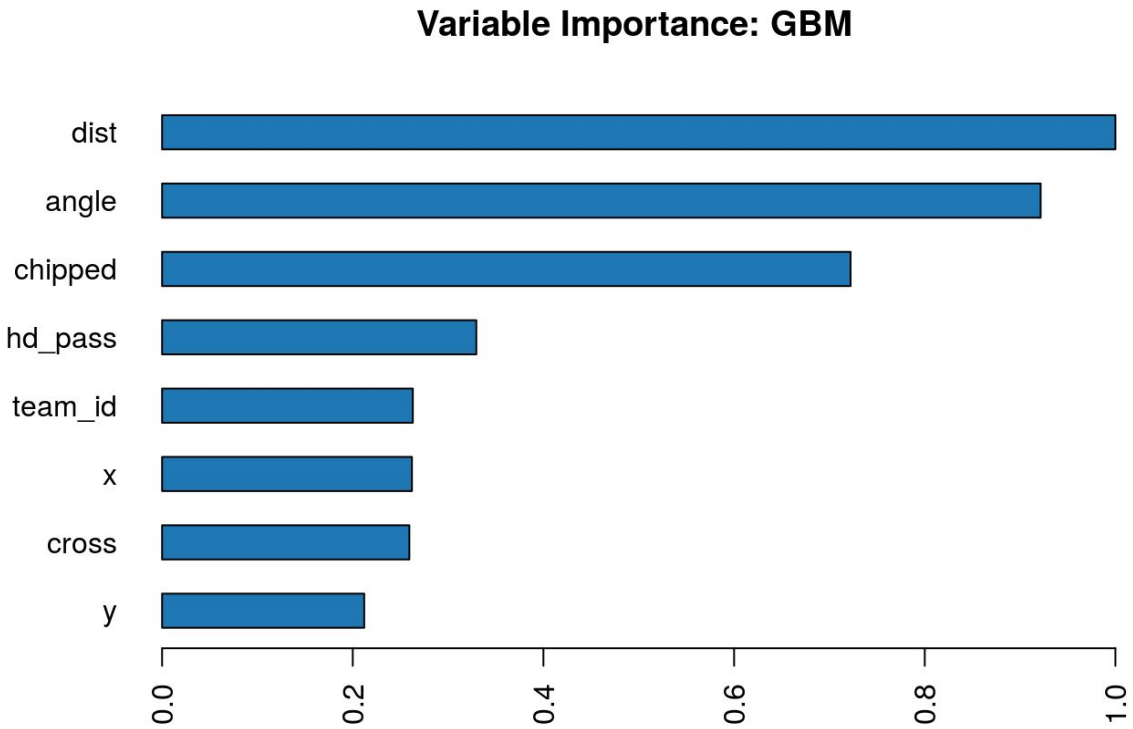
@mjskay

boB Rudis (hrbrmstr)

rud.is/books/creating-ggplot2-extensions/

Future directions

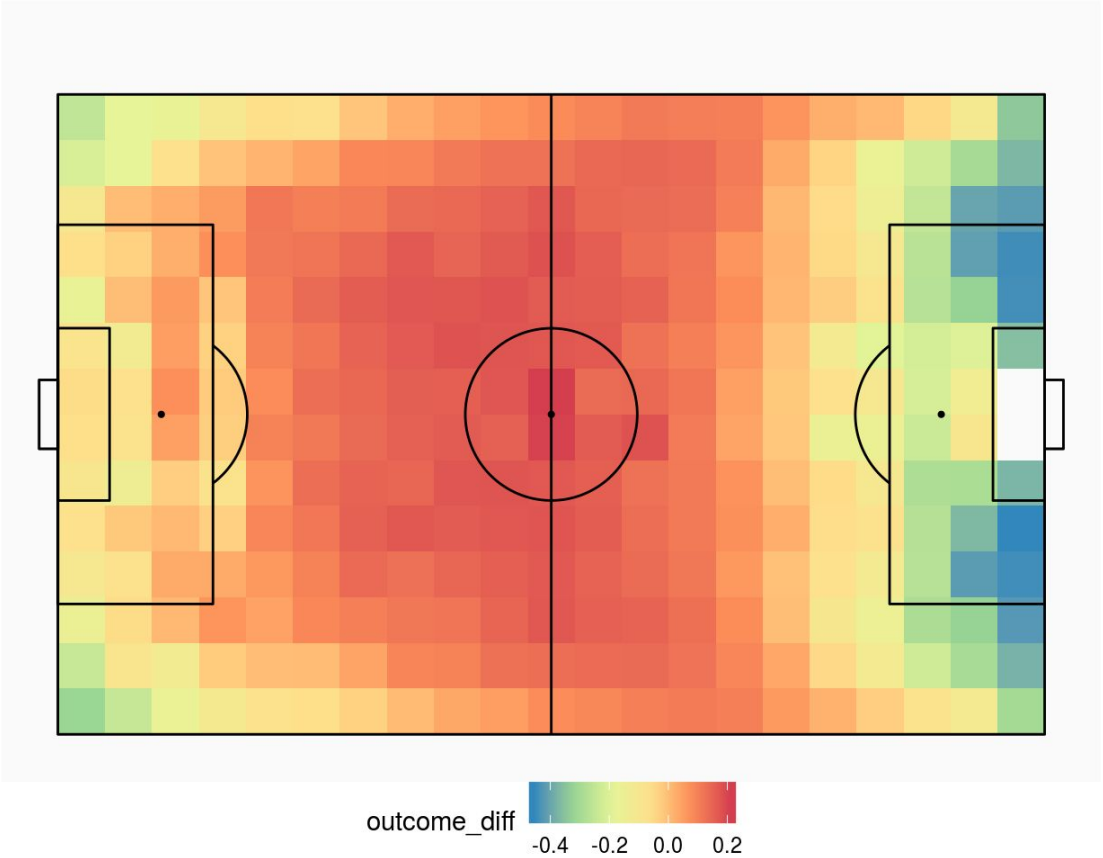
Machine learning



caret

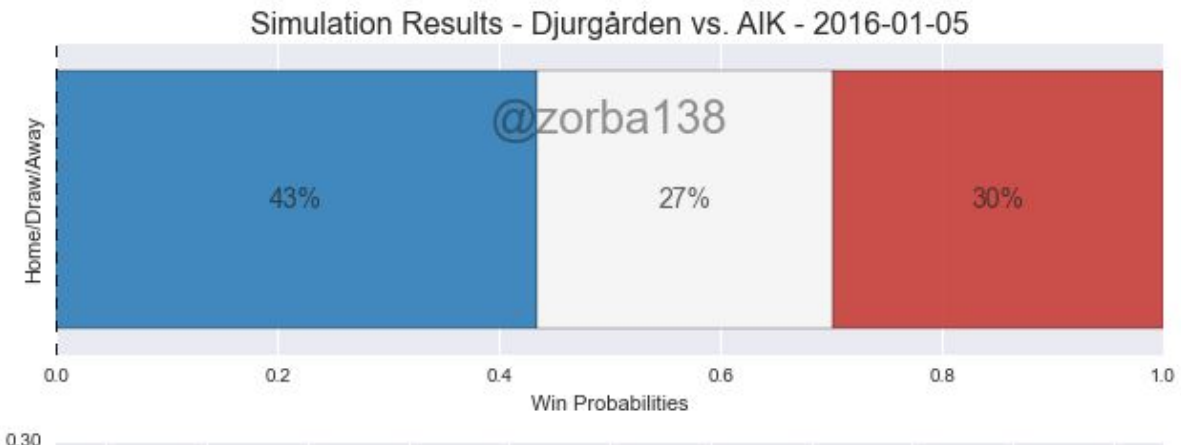
Future directions

Machine learning xPass



Future directions

Match simulation



Future directions

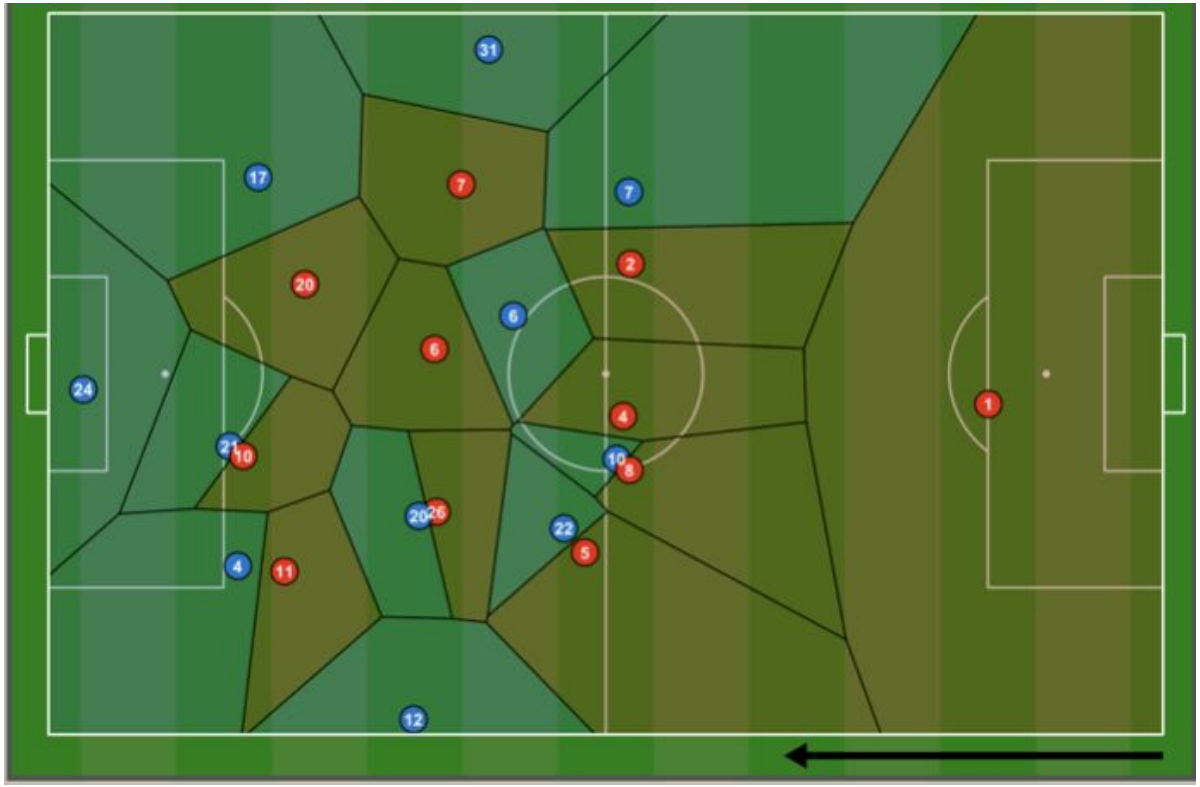
xPts

№	Team	PTS	xG	xGA	xPTS
1	Manchester City	80	82.06 -0.94	22.48 +1.48	77.25 -2.75
2	Liverpool	79	66.95 -5.05	24.09 +5.09	70.60 -8.40
3	Tottenham	64	53.45 -6.55	40.67 +6.67	53.35 -10.65
4	Arsenal	63	54.01 -10.99	42.62 +3.62	50.67 -12.33
5	Chelsea	63	53.70 -1.30	32.80 -1.20	59.76 -3.24
6	Manchester United	61	60.24 -0.76	44.74 +1.74	53.34 -7.66
7	Wolverhampton Wanderers	47	43.41 +3.41	34.72 -4.28	49.84 +2.84

understat.com

Future directions

Space control



Voronoi tessellation

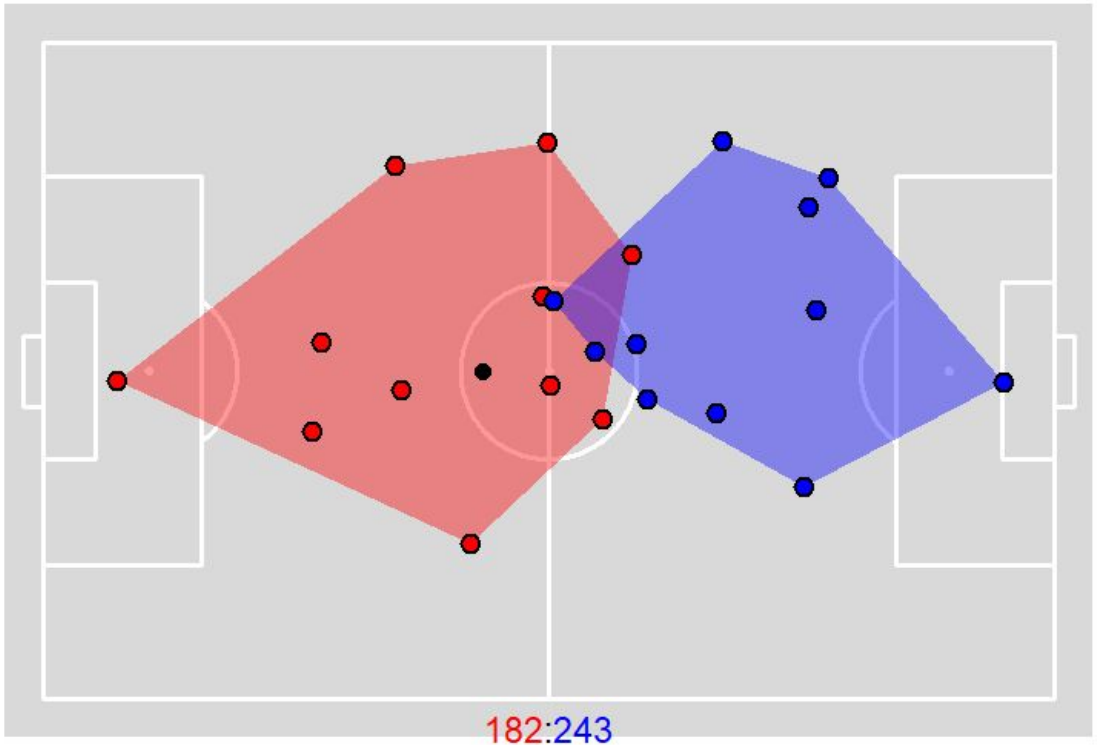
sp

deldir

@JoeyHarings

Future directions

Space control



Convex hulls

gganimate

tweenR

Thank you!

jogall/soccermatics

@joedgallagher