

Automating insight extraction from football data visualizations

Daniel Girela / Twenty3 Sport / daniel.girela@twenty3.sport

- 1

Football data visualizations convey an incredible amount of information very quickly. Moves, passing directions and, virtually, every type of event that can be recorded in a football match is able to be represented on a graph.
- 2

A simple method to extract insight from a certain type of visualization is to execute it over two different datasets (events by two different players, or just by a player when playing in two different positions) and compare the outputs.
- 3

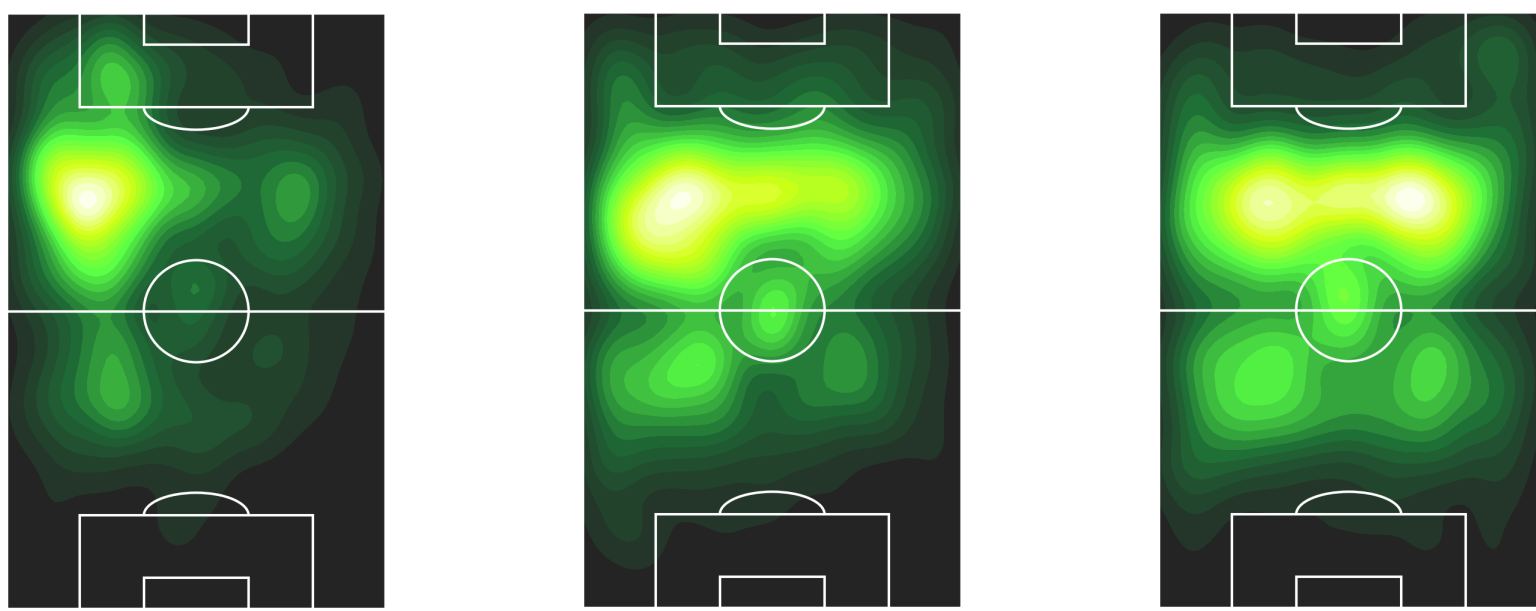
We define metrics that compare visualizations (heatmaps & passing sonars) to automate this process of extracting insight.
- 4

Everything below is built on Wyscout event data for the last two Premier League seasons.

Heatmap Comparisons

A heatmap is a measure on the pitch.

We compare two heatmaps by measuring the cost of optimally transporting one to the other (this is the earth mover distance).



Silva

Maddison

Eriksen

David Silva's, James Maddison's and Christian Eriksen's heatmaps in EPL 2018/19

emd(Maddison,Silva) = 0.0034 emd(Silva,Eriksen) = 0.0087 emd(Maddison,Eriksen) = 0.0019

Passing Sonar Comparisons

A passing sonar is a 16-dimensional vector with positive components that add up to 1.

We compare two passing sonars via a modified l¹ distance that penalises differences in distant components more than close-by ones.



Sterling

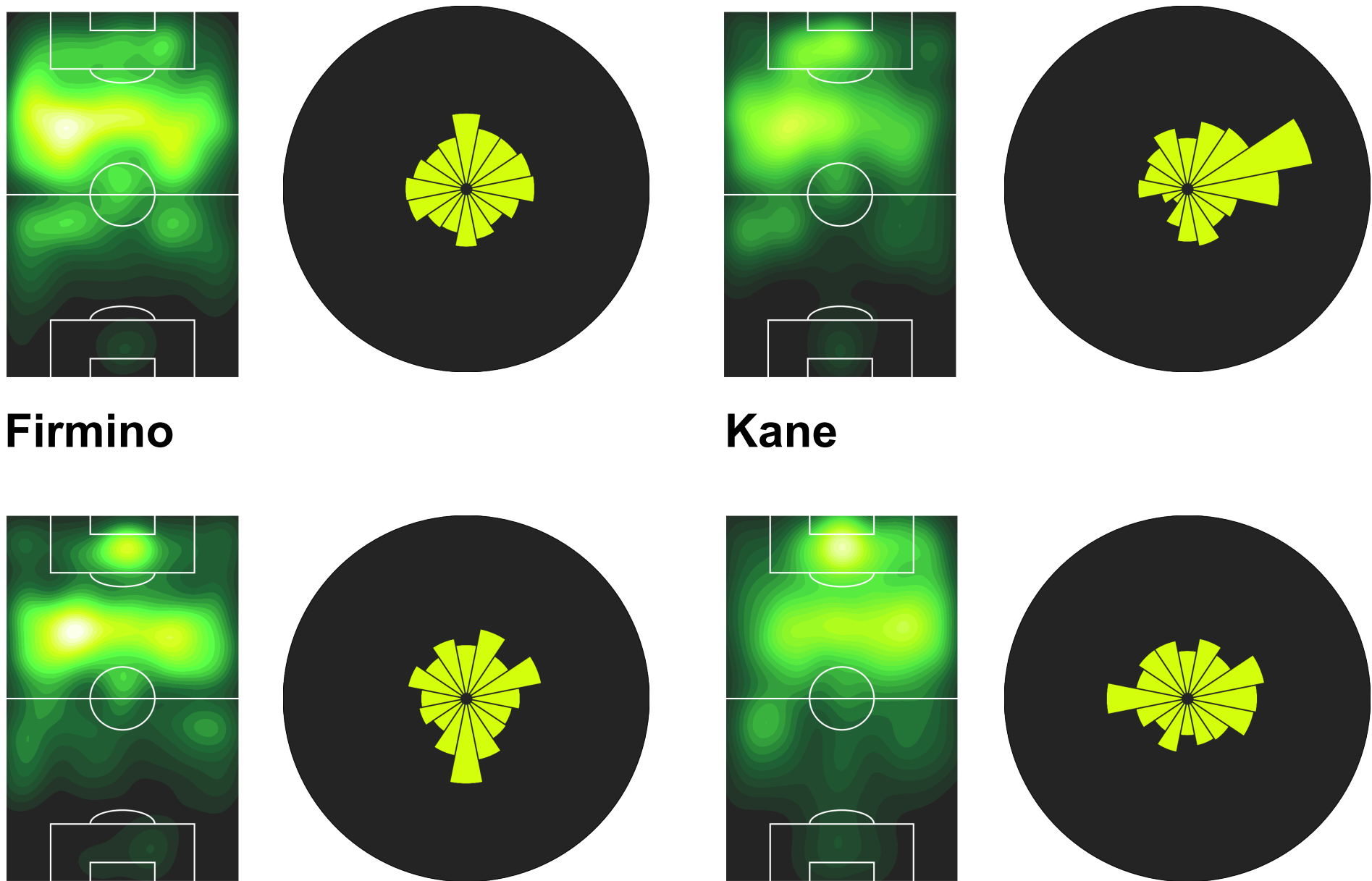
Hazard

Mané

Sterling's, Hazard's and Mané's passing sonars of all their passes in EPL 2018/19

d(Sterling,Hazard) = 3.565 d(Sterling,Mané) = 2.746 d(Mané,Hazard) = 1.314

Bivariate Comparisons



Firmino

Kane

Rondón

Wilson

Firmino's (top-left), Kane's (top-right), Rondón's (bottom-left) and Wilson's (bottom-right) heatmaps and passing sonars, respectively.

Compare Firmino to other EPL strikers with, at least, 2000 minutes on field last season

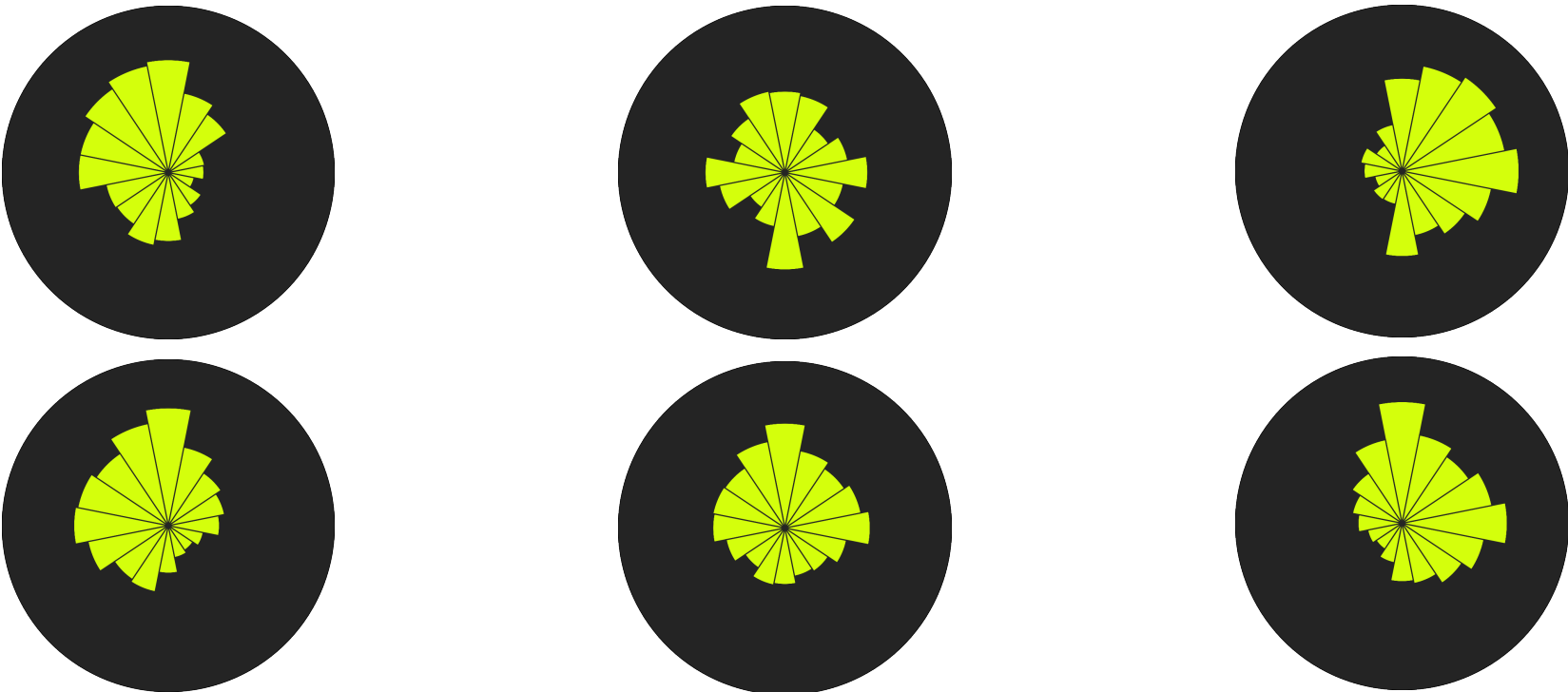
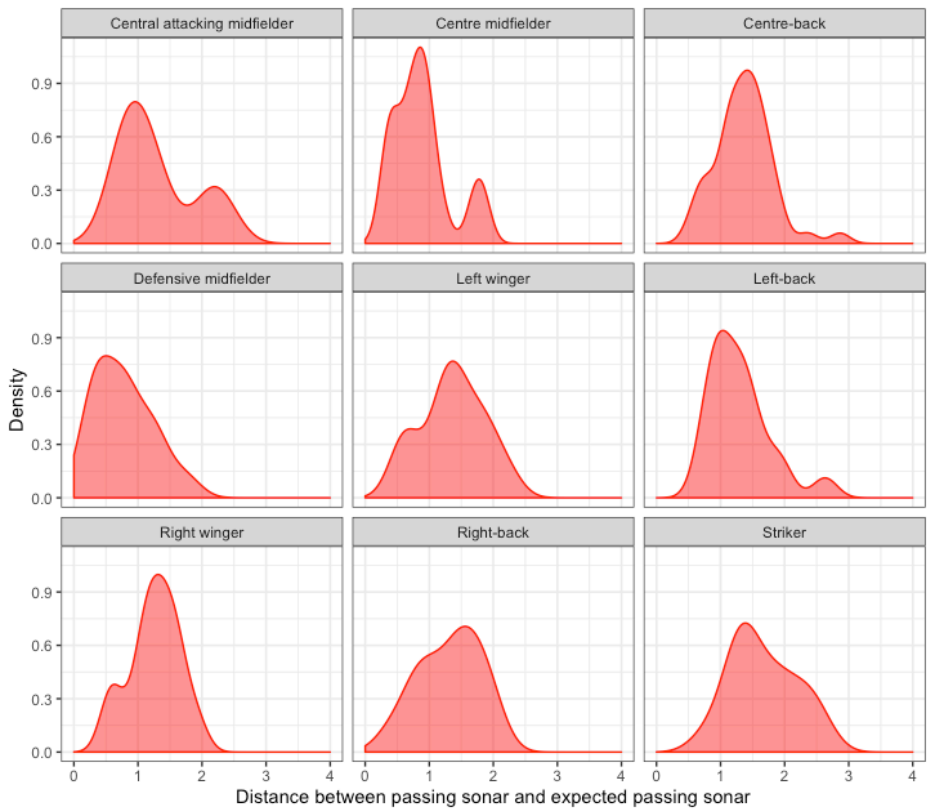
	Rank based on distance to Firmino's heatmap	Rank based on distance to Firmino's passing sonar
Kane	2	21
Rondón	1	3
Wilson	16	1

Expected Passing Sonars

The expected passing sonar of a player given his heatmap is the passing sonar an average player would have if his heatmap coincided with the heatmap of the player under study.

- First, discretize the pitch into 16 x 12 square cells
- Then, compute the average passing sonar in each cell
- Finally, compute an average of the cell-passing sonars, weighted by the heatmap mass in each cell

It is robust: distances between passing sonars and expected passing sonars are generally small. It serves us to classify players in terms of their passing predictability.



High predictability:
David Brooks

Average predictability:
Richarlison

Low predictability:
Felipe Anderson

(Top: Real passing sonar; bottom: expected passing sonar given the player's heatmap)