

Goalkeeper Passing- What's been done so far

By Salil Akundi • 27 Jan 2019 • 8 min read

Background

It is true that a goalkeeper's main job is to keep the ball out of the net. However, in the modern game, where margins are much tighter and errors are easily punished, it is a valuable advantage for a goalkeeper to be an adept passer of the ball. This is particularly useful when a team plays a style with the intent of creating spaces to attack by drawing the opposition to press high, such as the style employed (with success, in the past with Napoli) by Chelsea manager Maurizio Sarri. It is no coincidence, either, to see big clubs spending record sums of money on goalkeepers these days, goalkeepers that are not just shot-stoppers, but also excellent footballers.



Napoli playing out from the back to create one-on-ones higher up the field. Note how deep Napoli's center-backs are in the build-up and how the goalkeeper Pepe Reina shows composure under pressure to pass the ball instead of simply lumping it upfield.

It's more than just the goalkeeper

With this in mind, we thought it might be handy to develop a tool that determines how good of a footballer a goalkeeper is. It is too simplistic to say that goalkeepers with better passing success rates are better at passing than those with lower rates. This is because of two reasons: firstly, there are lots of other factors at play such as the play style of the team (building out from the back vs direct play), quality of the opponents in pressing high up the pitch, the starting formation of the team (for example, 3-5-2 vs 4-4-2 vs 4-3-3) among others.

Secondly, the success rate of a pass depends on multiple factors- the ability of the passer, the awareness of the receiver to adjust his angle and body shape to receive the pass and the general positioning of the team. In addition, the completion of a pass per se cannot classify it as a 'good' pass. For example, the weight of the pass from the passer could be too weak or too strong, which would not allow the receiver to take a good touch and could lead to loss of possession. One possible way of accounting for this would be to determine how many passes were completed after the first pass and weighting subsequent passes accordingly. However, the issue with this is that a loss of possession in turn could be due to a bad decision of the passer or a mistake by the receiver or neither- in the case of very good pressing or a poor structure of positional play.

The clip below taken from a match between Liverpool and Leicester City in 2017 shows one example of poor positional play and spatial awareness. When the ball is played from the goalkeeper to the defender, the defender has no real options for a pass except for a pass back to his own goalkeeper. He instead takes the risk to pass it forward, expecting his midfielder to move into the space. However, the midfielder is too late to arrive, allowing Leicester City to capitalize and score.

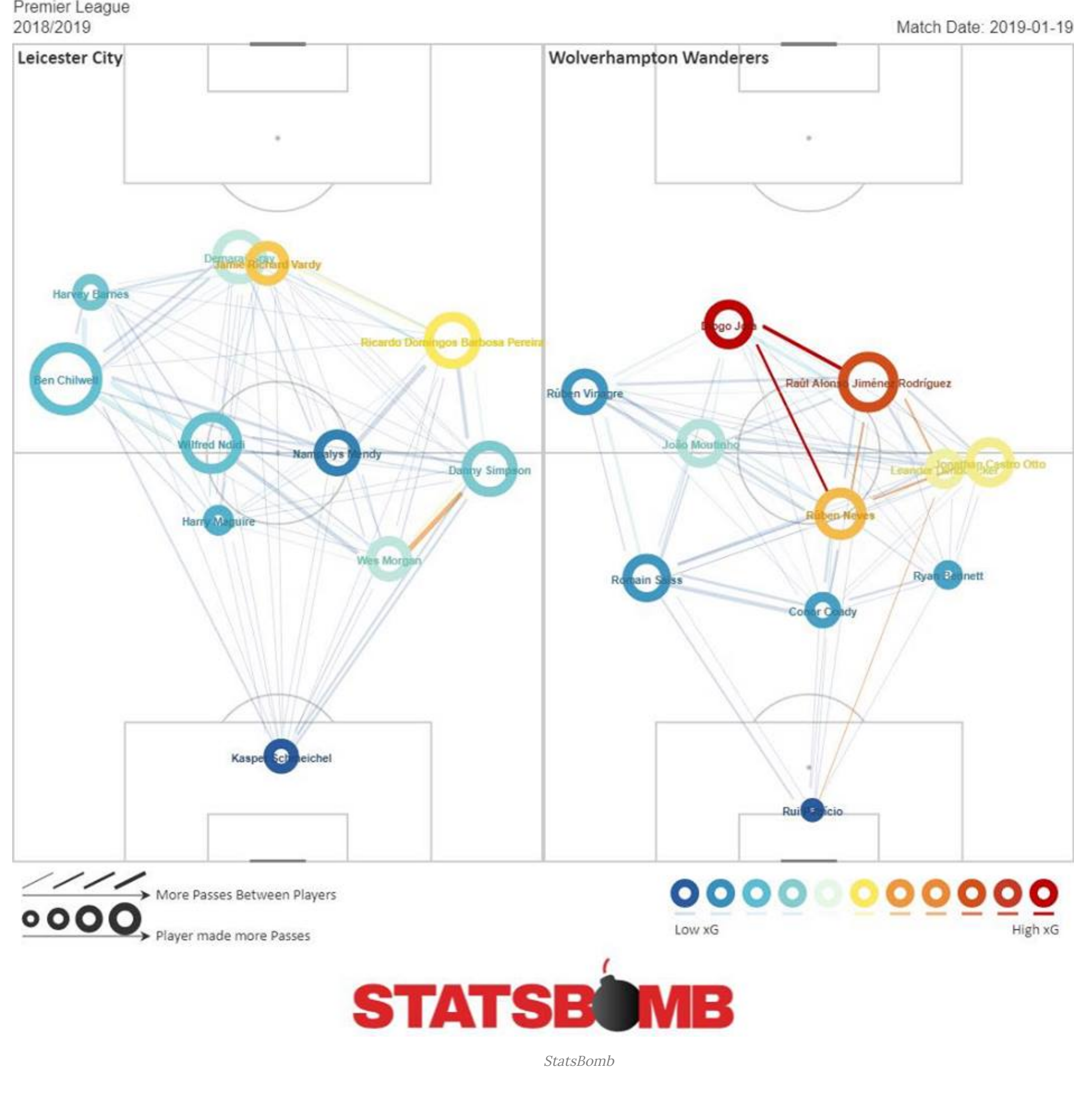


via GIPHY

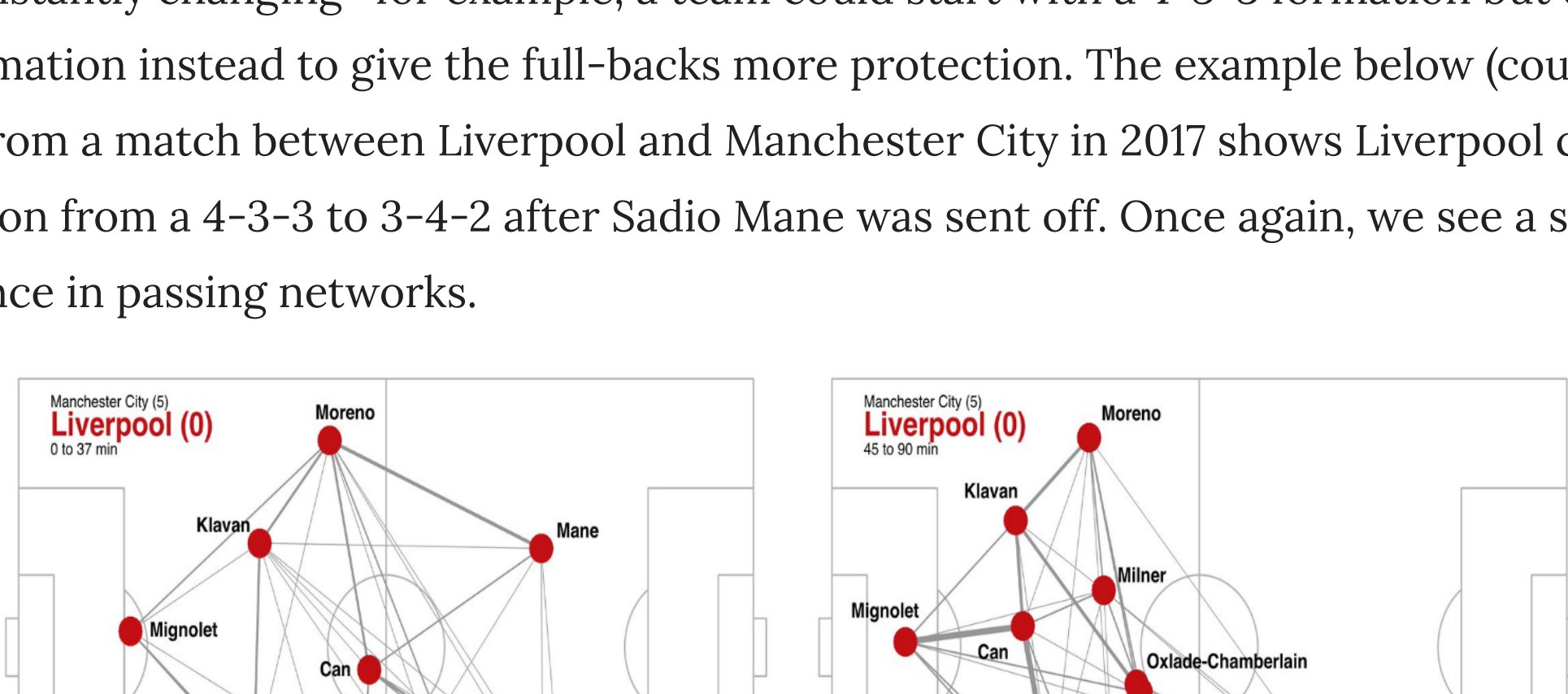
The tool we propose takes the aforementioned factors into account and attempts to remove any bias that may occur as a result of those factors. This is a challenge because soccer, more so than other sports, is a chain of events, with decisions often being made based on the game state. For example, a team leading 1-0 in the 85th minute of a game is likely to play less expansively than if the same team were winning 3-0 at that point.

Often forgotten is the fact that in sport, we are dealing with human beings and not machines. Like most human beings, players are prone to dips in confidence when things go against them. It is often seen that when teams are on a losing streak, they tend to play more conservatively. Likewise, when teams are on a winning streak, not only do they take more risks, but their passing also becomes quicker and sharper. In our opinion, our tool should reward those players that maintain a consistent level of passing, regardless of how the team as a whole performs.

Let's have a closer look at how the formation of a team can affect the passing options a GK has.

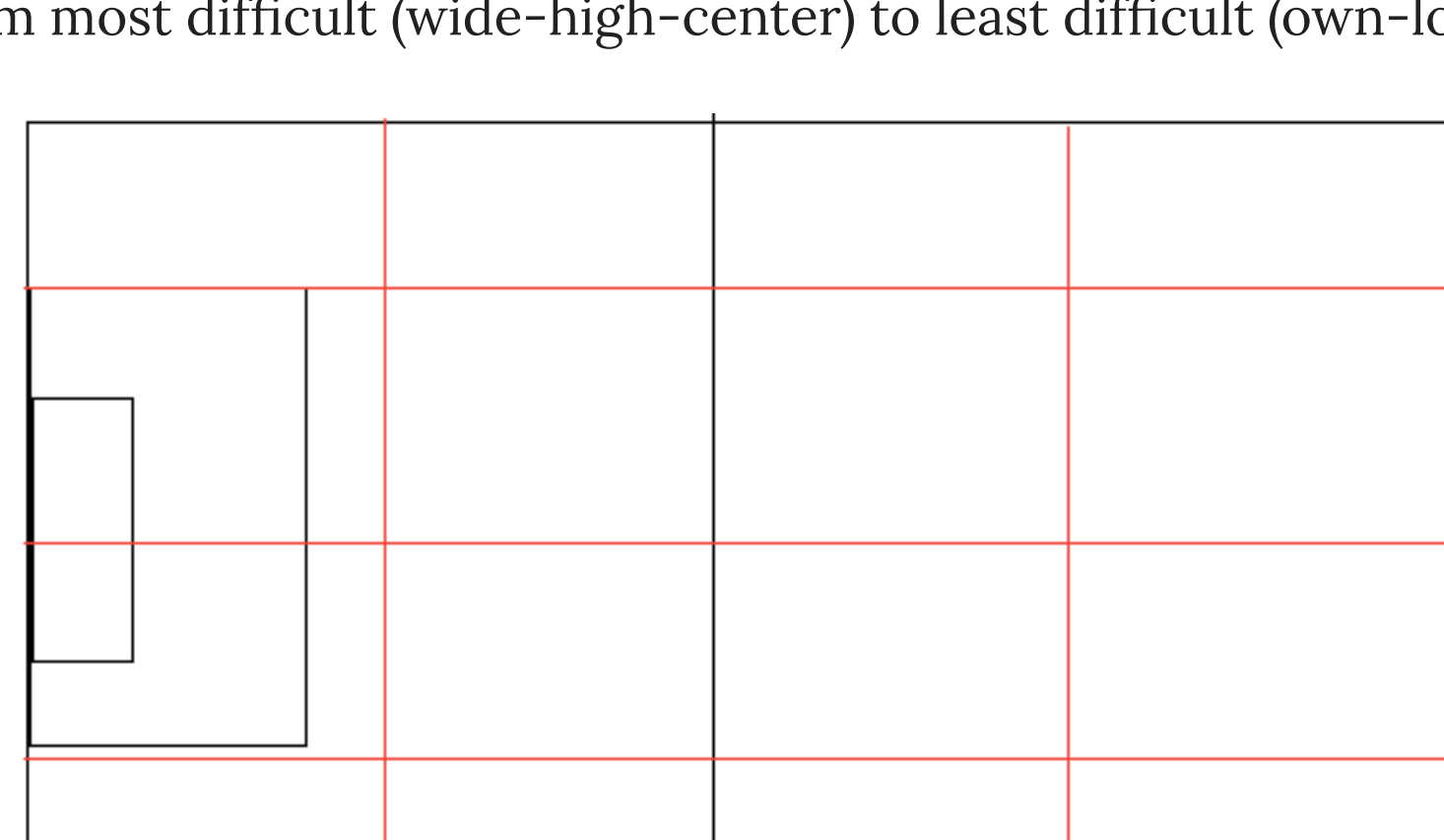


As a simple example, the picture above (courtesy of StatsBomb) shows how the passing options a goalkeeper has is determined by the starting formation of the team. In this particular match, Leicester City played in a 4-2-3-1 starting formation, while Wolverhampton Wanderers played in a 3-4-1-2 formation instead to give the full-backs more protection. The example below (courtesy of Opta) from a match between Liverpool and Manchester City in 2017 shows Liverpool changing formation from a 4-3-3 to 3-4-2 after Sadio Mane was sent off. Once again, we see a similar difference in passing networks.



When the case of Arsenal (who played 33% of their games this season with 3-4-3) was considered, a similar pattern was found wherein the passing networks (too messy to show here!) showed thicker lines from the goalkeeper to the center-backs when a 3-4-3 formation was used.

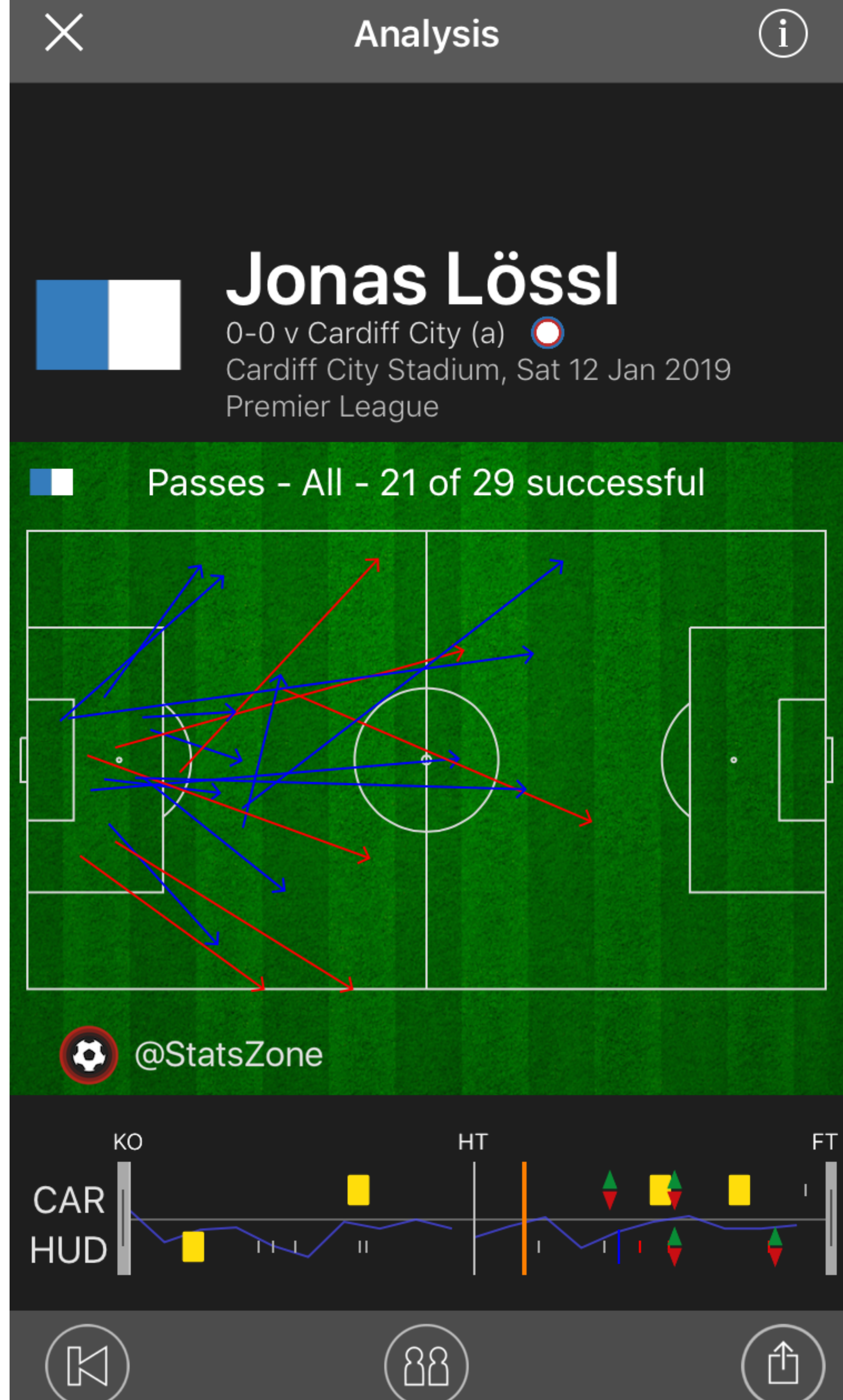
Next, we take a look at 'distance and direction' of a pass. Most sources of statistics only differentiate between 'short' and 'long' passes. However, our research shows that pass accuracy fluctuates widely based on direction as well. This also has to do with the fact that some goalkeepers in the league are naturally right-footed, while others are left-footed. Moreover, some goalkeepers, such as Alisson, prefer to throw the ball out more, whereas others such as Kepa, prefer to use their feet. To make a distinction, we decided to split the pitch into 8 different zones, as demarcated by the red lines shown below. Passing networks are then looked at on a game-by-game basis to determine the 'weaker' side and passes into each of these 8 zones are weighted accordingly, from most difficult (wide-high-center) to least difficult (own-low-center).



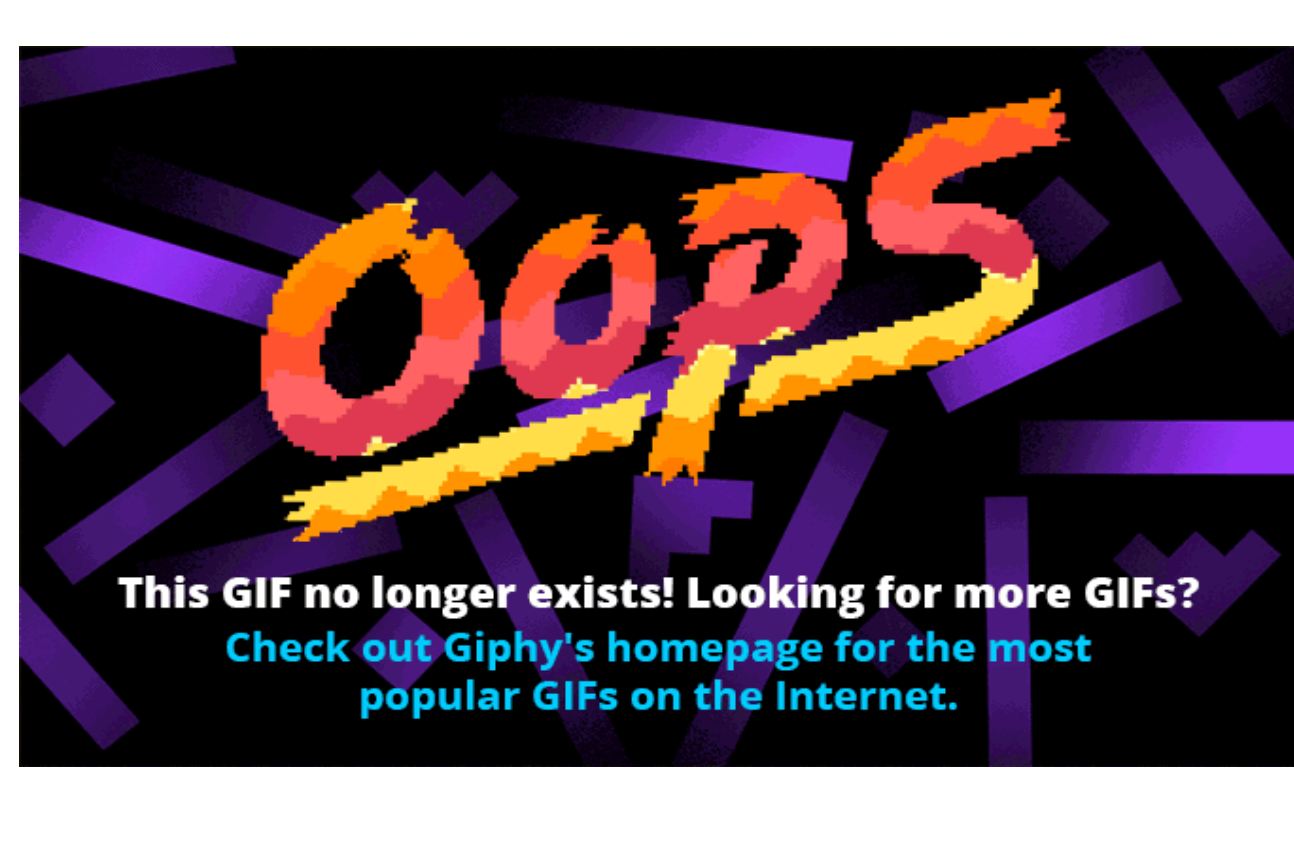
Other Issues and Future Plans

One of the issues we faced was the lack of available data. While commonly used metrics such as passing percentages and starting formations are readily available, others such as ways to measure 'pressure' or speed of distribution are not. (We note here that more extensive data is available solely for the FIFA World Cup 2018. However, we decided against using our model for the World Cup since it is well accepted within the football community that teams in an international tournament do not have as developed a style of play compared to club teams due to lack of preparation time. Moreover, the sample size of games is extremely small.)

In addition, as the picture below shows, most sources of data do not differentiate between passes that are intended towards a specific player and hopeful punts up the pitch. This meant that we had to review video footage of games in order to determine these ourselves. Fortunately, the StatsZone app provides a slider which lets us determine each minute a goalkeeper passed the ball.

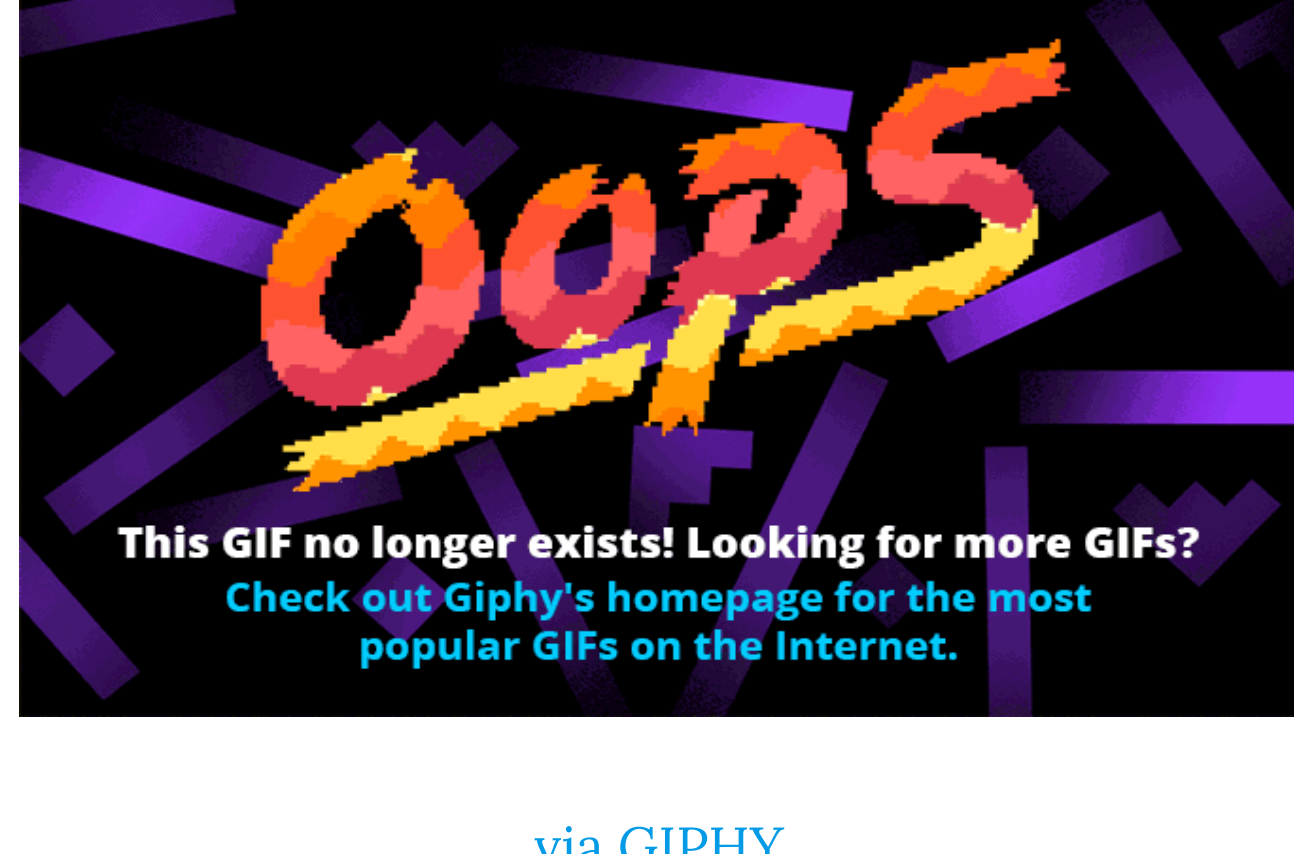


The following examples will illustrate the difficulty of distinguishing between a pass and a punt just from numbers. In this clip taken from a match between Huddersfield Town and Manchester City in 2018, we see that all Manchester City players, except for the striker, Sergio Aguero are either being man-marked or put under pressure. Ederson, the goalkeeper, then thinks quickly plays a pass specifically intended for Aguero, who then scores.



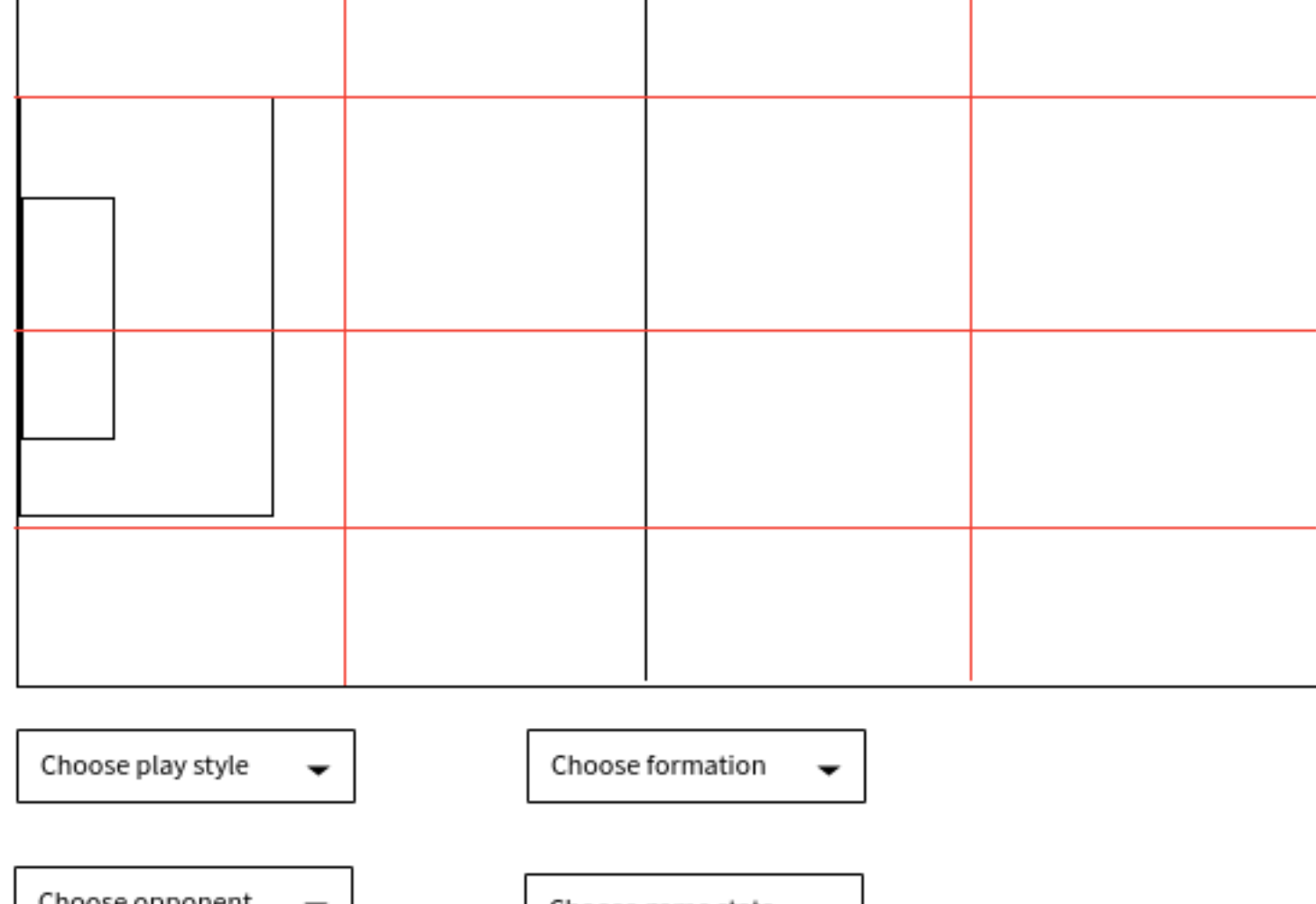
via GIPHY

This clip, taken from a match between West Ham United and Burnley, shows an example of a hopeful punt upfield from Joe Hart, as he is being put under pressure by a Burnley attacker. Hart gets given the assist, but it is fair to say that he did not really mean it!



via GIPHY

To help us solve this (and other issues), we developed a little app using R Shiny (screenshot below). Punts and passes are color-coded differently, but both are retained since we would like to use 'punts up the pitch' to help us determine the goalkeeper best suited to a 'long-ball team'. The tool also helps us categorize passes by zones of the pitch that they are played into, as described earlier and calculates a metric weighted based on the difficulty of a pass into a particular zone and the opponent/playing style being used. At present, the weights we have chosen are somewhat arbitrary. However, in the future, we plan to come up with an unbiased overall metric for goalkeeper distribution based on empirical weights. More on this will be described in a blog-post that will be published in March.



Sources and Credits

We would like to credit a number of sources. StatsBomb, Stats Zone and Opta were used for passing networks and visualizations. The mock pitch tool was inspired by Ben Torvaney's 'soccer event tracker'. Full match videos from the 2016-17 soccer season onwards (this would probably interest a lot of people!) were found on fullmatchsports.com/fullmatch.

Please feel free to write to me at salil1999@gmail.com if you have any suggestions or ideas on other metrics we could add to our model or any thoughts in general!