**StatsBombR Analysis of Hertha BSC-Mainz 05 (12/20/2015)**



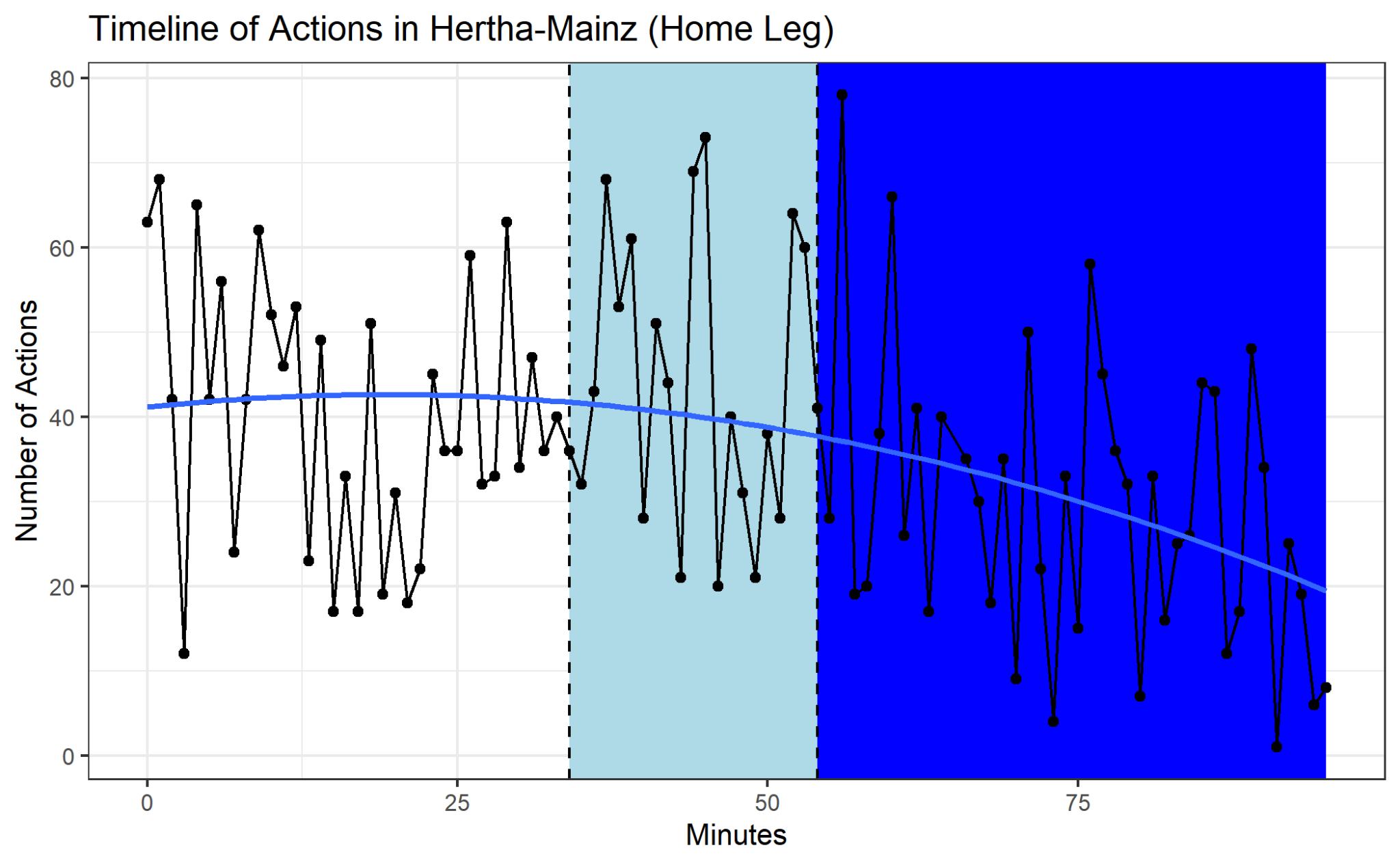
Theo Schmidt

D.C. United Data Analytics Internship Technical Project

04/07/2024

In examining the StatsBomb data, I quickly realized that free data was not available for the majority of soccer leagues over a long term time span. The competition data included a variety of leagues and competitions, both men’s and women’s, but having grown up in a German-American dual-national family, I tend to gravitate towards the Bundesliga. The only season available for Bundesliga data was the 2015-16 season, in which my favorite club, Hertha Berlin, finished 7th and qualified for the Europa League. With this data, I decided to analyze one of Hertha’s games, and randomly selected the final game of the calendar year, a late December home match against Mainz.

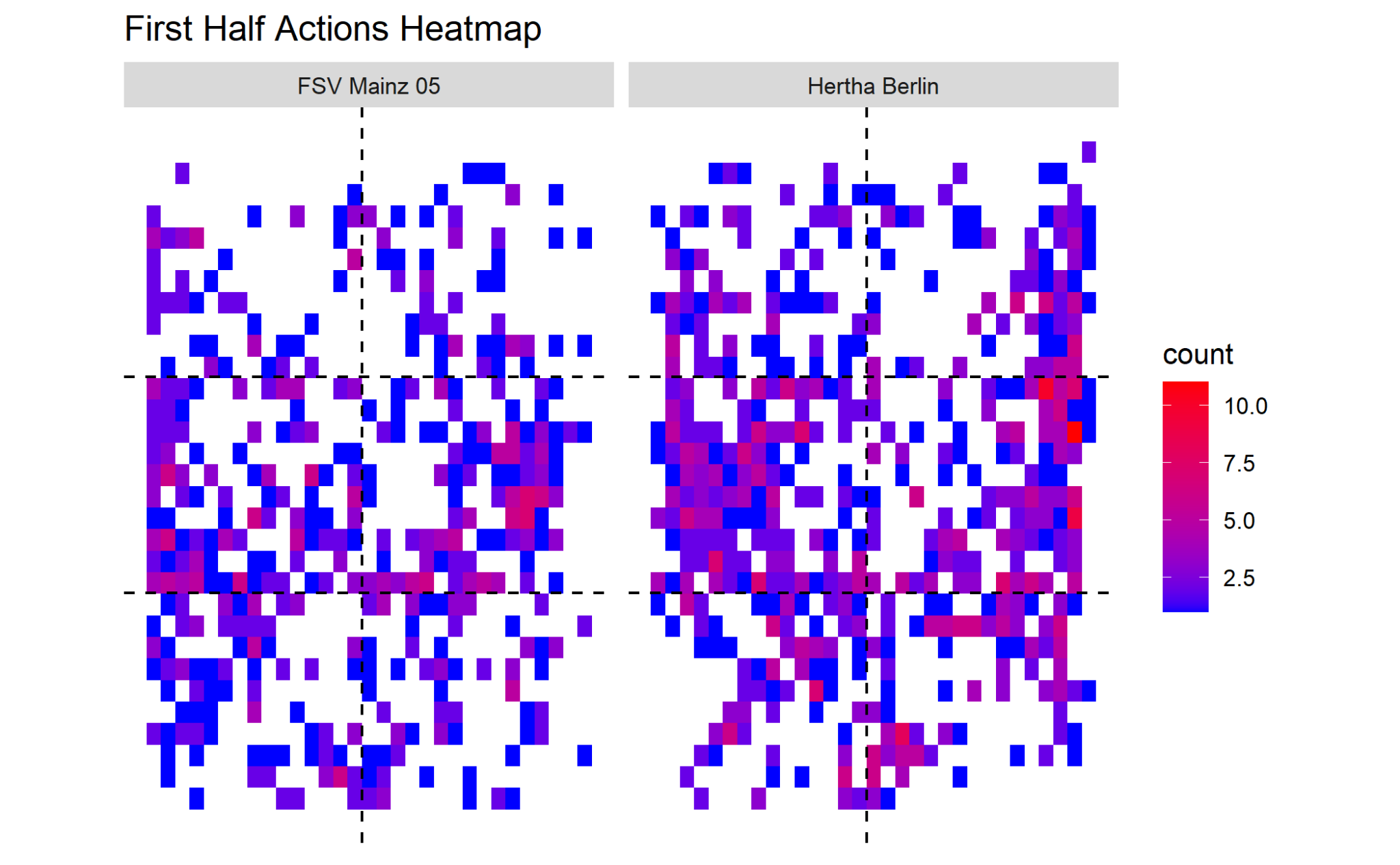
In getting the proper data from R, I first had to download the free competitions package to identify the season and competition IDs. From there, I located the JSON file for the game in question and loaded it into R. Game flow in soccer often changes along with the scoreline, and in my time keeping statistics for the Syracuse University men’s soccer team, I’ve noticed the structure of the game can break down as the game wears on. To see if this same phenomenon exists at a high level of the professional game, I wanted to create a simple line plot documenting the quantity of actions per minute across the game.



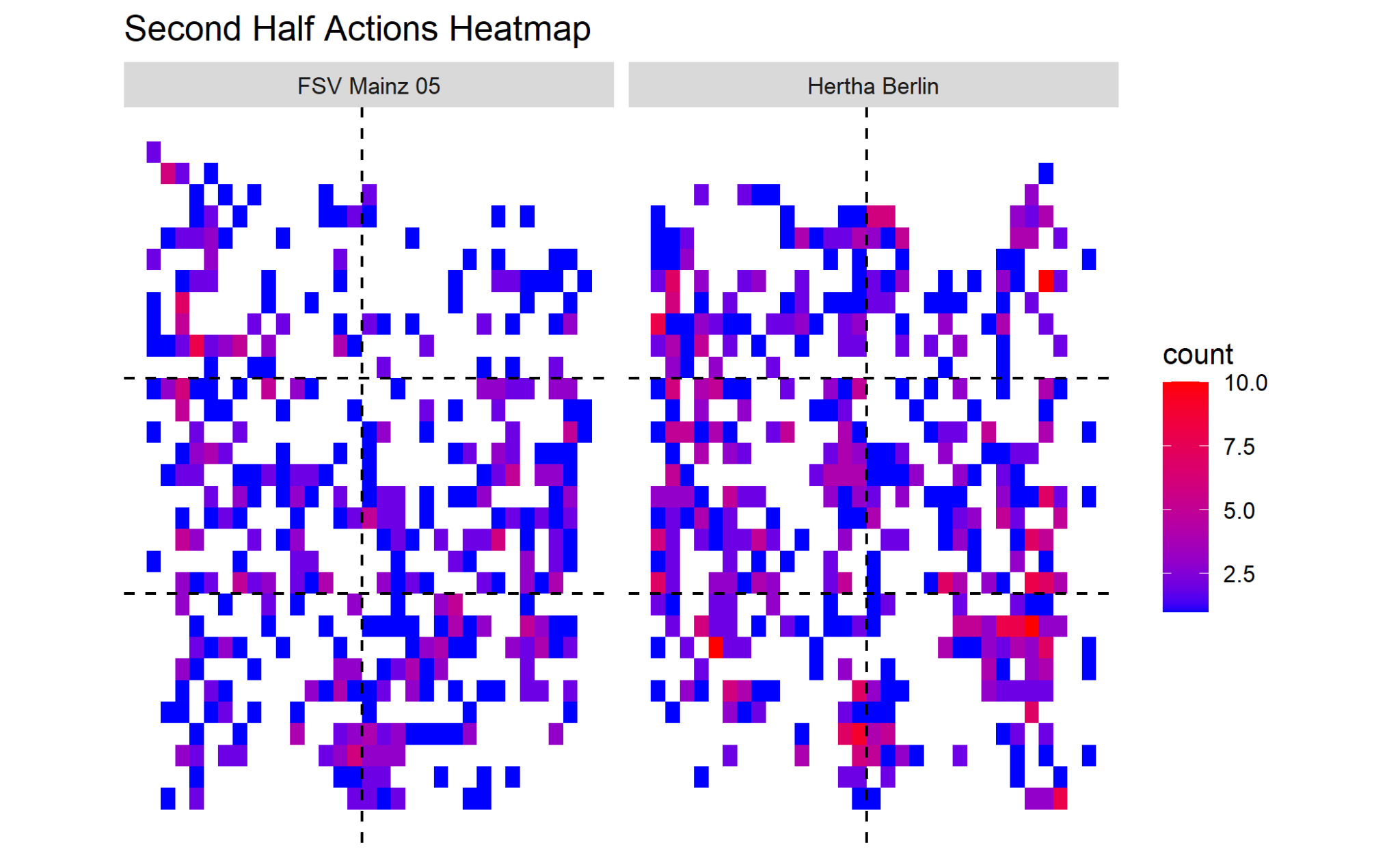
The plot above depicts the quantity of actions per minute across the Hertha-Mainz game. The plot is broken up into three sections; one from kickoff to the 34th minute, a second from the 34th to 54th minute, and a final one from the 54th minute to the final whistle. These sections depict the scorelines, starting from 0:0 and changing colors when Hertha’s Vladimir Darida opened the scoring 34 minutes in. The second bar shows the minutes in which Hertha held a 1:0 lead and the third reflects the eventual 2:0 scoreline brought through Salomon Kalou’s 54th minute goal. The quantity of actions per minute remained relatively constant for the first half of the game, however decreased steadily in the second half.

The point at which actions per minute really start to decline unsurprisingly coincides with Kalou’s goal. At this moment, Hertha has a steady lead and can afford to take a slower, more defensive approach. For Mainz, the approach becomes the reverse as they fight to pull back two goals within the final half hour of the game. The blue quadratic line was inserted to depict more clearly how the average number of actions falls steadily throughout the second half. A decrease in actions and productivity should also be an indicator to coaches that their players are tiring. The point around minute 60 when the actions really start to taper off is also a common point for coaches to begin substitutions, and the data seems to confirm this approach.

From there, I wanted to make use of the tracking data provided by the StatsBomb database. We can get a better feel for where on the pitch each team is more active to see where Hertha was able to exploit and win the game.



This heat map depicts total actions for each team in the first half. The majority of action came in the midfield, particularly on the left for Hertha. One of Hertha’s top players in this season was Kalou, their left midfielder and a former champions league winner. Additionally, left back Marvin Plattenhardt solidified himself as a centerpiece in the squad, and even received call ups to the German national team in future years. The left side of Hertha’s squad contained two of their top players, which is a possible reason for the increased density on that half of the pitch. Hertha was also particularly active on the right sideline, although without as consistent of density as on the left side. Mainz by comparison appear to have been less active on the wings.



The second half was not as telling for either team, which makes sense when we compare our heatmap for the half to the earlier line plot. Remember that the second half saw a rapid decrease in actions per minute, particularly after Kalou extended Hertha’s lead and put Mainz somewhat out of reach. In the game, Hertha failed to make its first substitution until the 77th minute, which is rather late. It would not be hard to see how the fatigue of the players brought about by late substitutions contributed to fewer actions as the game dragged on.

Data such as content found in the StatsBomb library can be useful for professional clubs both in-game and postgame. By analyzing action frequency and locations throughout a game, coaches can manage their rosters more effectively. Crucially, since that 2015 match, the quantity of substitutions allocated in a Bundesliga match has increased from three to five per team. This rule change has added a more significant strategic element to in game substitution, which can be aided by data analysis. Although Hertha still won the game, they performed much less efficiently in the final 30 minutes, and an earlier substitution influenced by team/player heatmaps and team output data could have allowed them to keep their foot on the gas.