vis-01

Silke Meiner, s81320

June 2020

Abstract

We visualize performance scores for soccer players at different field positions. We Show a series of two-dimensional scatterplots for players at three German clubs that participate in Bundesliga. The plots are based on SOM plots trained for all players in Bundesliga and the attributes shown were bundeled from a variety of special but highly correlated attributes.

1 Introduction

We work with the soccer data from FIFA / kaggle. It contains data on about 18.000 players under contract with FIFA associated clubs. Our data set has 80 attributes, 26 of which give scores (range 0 to 100) on distinct field positions and a preferred position. We focus on players associated with clubs in the German premier league (Bundesliga). We want to find an appealing visualisation.

2 Dimensionality reduction

We expect players to perform similarly in positions that are close in the sense of figure 1. How exactly is not clear.

Figure 1: Soccer field positions abbreviations

We run a correlation analysis on the scores for the field positions. Figure 2 shows the correlation matrix. Attributes are row wise positions in figure 1.

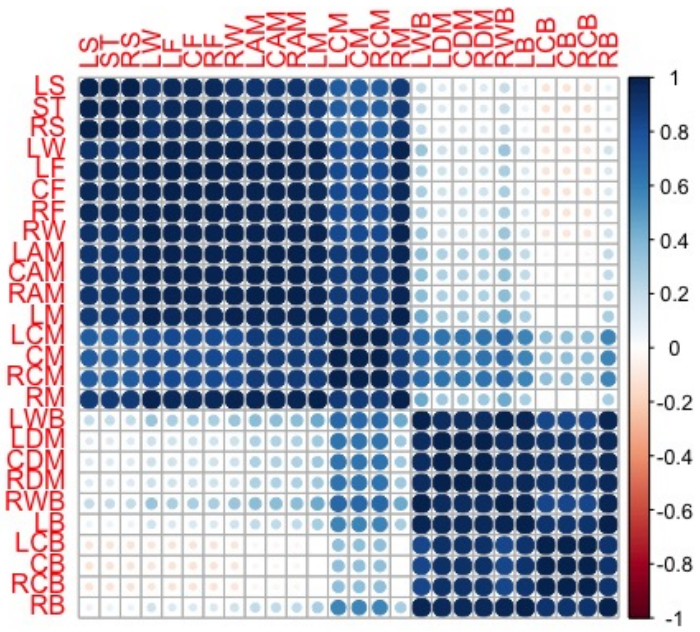


Figure 2: correlation of scores

We observe that the topology of field positions as shown in figure 1 agrees with the correlation in figure 2. Note that the new grouping is not exactly row wise, as seen in positions LM and LB on the extreme left and RM and RB on the right. We group positions according to correlation:

- offense: LS, ST,RS,LF,CF,RF,RW,LAM,CAM,RAM, LM, RM
- center: LCM, CM, RCM
- defense: LWB,LDM,CDM,RDM,RWB, LB, RB
- back: LCB, CB, RCB

We calculate new performance scores as the maximum of any original position in the newly formed group of positions. The correlation we obtain now captures the previous correlation with less attributes, cf figure 3.

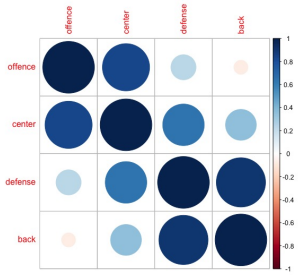


Figure 3: correlation of scores, grouping: offense, center, defense, back

3 German premier league, Bundesliga

We will now focus on clubs from Bundesliga in its 56th season 2018/19. Correlation of original and grouped attributes will look very similar to figures 2 and 3. We will work with the same grouping of field positions as in section 2: offense, center, defense and back.

4 SOM

We present a SOM for the clubs participating in the 56th Bundesliga.

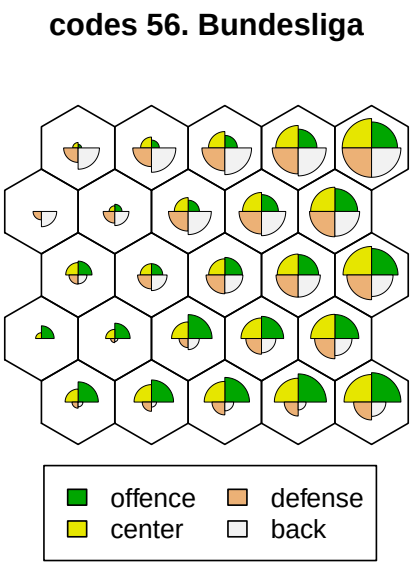


Figure 4: SOM codes for Bundesliga clubs

The SOM places out-performing players in the top right corner, players with high scores in offence and center on the right border with descending scores for defense and back (top to bottom in right border units). On the left border scores are low with the top prioritising back and defense and the bottom offence and center. The middle row represents players with scores uniform on all four attributes, ascending from left to right.

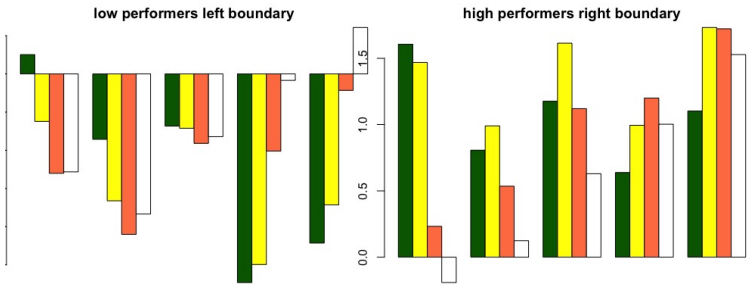


Figure 5: SOM codes at left and right boundary

As can be expected from high correlation between offence and center as well as between defence and back there are only few units with significant differences between the correlated atttributes. Traceable differences between defence and back can be followed in the top left corner in units 16,20 and 21. Differences between offence and center are traceable in the bottom left corner in units 1,6,7,8.

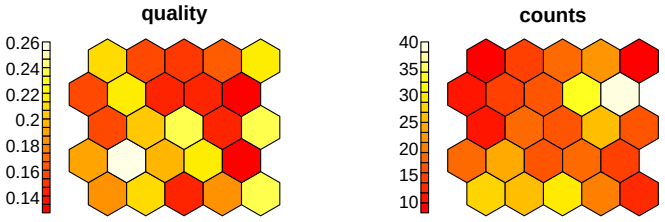


Figure 6: SOM codes for Bundesliga clubs

Quality is uniformly very good with even the high values only 0.26. The counts for players mapped to each unit places a high number of players in cells 19 and 20 (...) and low numbers in 5,10,15 and 25 (overall...).

4.1 clubs of Bundesliga in the SOM

We show a plot of 6 (of 18) clubs of Bundesliga and their players plotted in the SOM. The ordering is wrt the clubs final position: the first row is of clubs placing first, second and third (left to right) while the second is of clubs with the three bottom results.

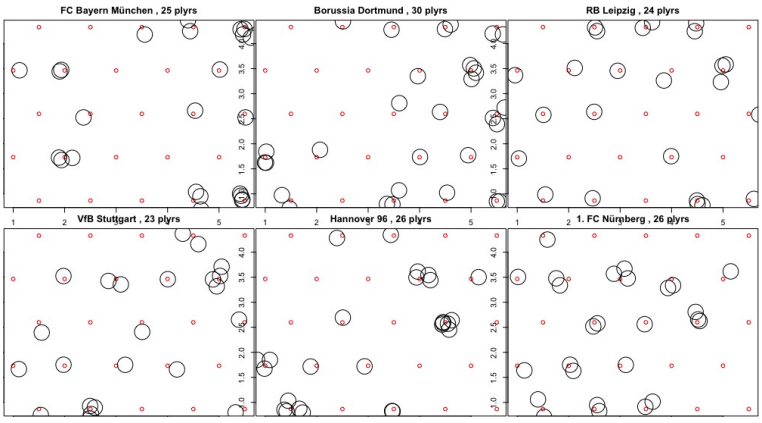


Figure 7: Each clubs players in Bundesliga SOM

We see Bayern Munich and Borussia Dortmund with many players at the right boundary, first placed Bayern Munich even more than Borussia Dortmund. Third placed RB Leipzig has fewer players at the right bound but some on the top or bottom where players are more dedicated to either defence, back (top row) or offence, center (bottom row). FC Nürnberg and Hanover 96 both have only a single player at the right bound. Otherwise they do not look too bad.

5 other layouts

Did we choose the best layout? As the difference b/w high and low performing clubs did not become obvious at a glance, maybe we can improve. We give plots for different layouts.

5.1 SOM, 2x3 layout

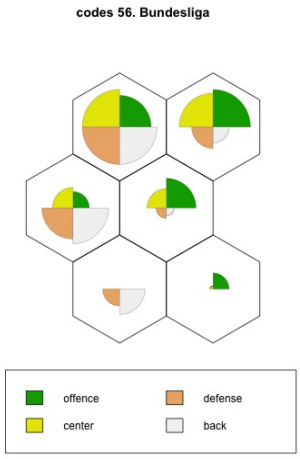


Figure 8: SOM 2-3 layout, codes

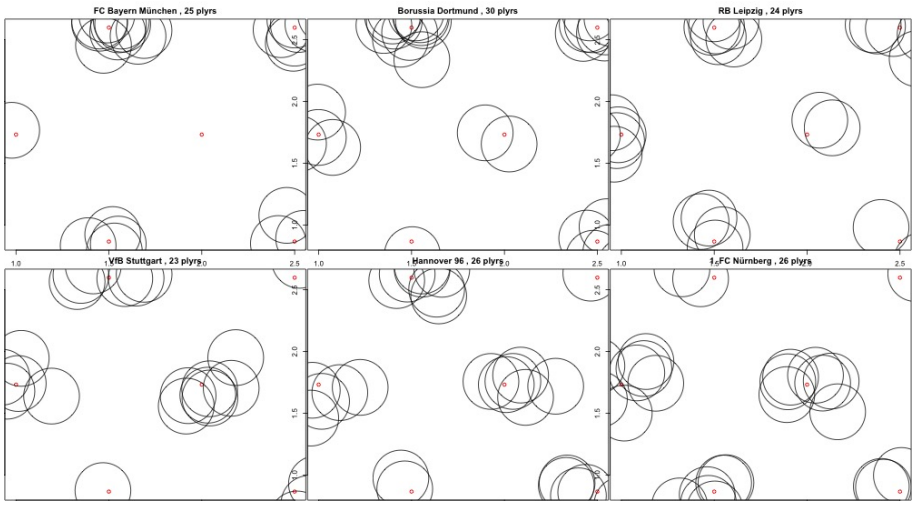


Figure 9: SOM 2-3 layout, clubs

5.2 SOM, 2x4 layout

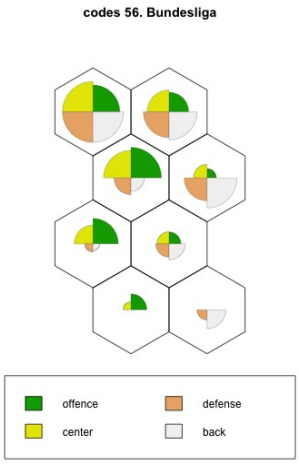


Figure 10: SOM 2-4 layout, codes

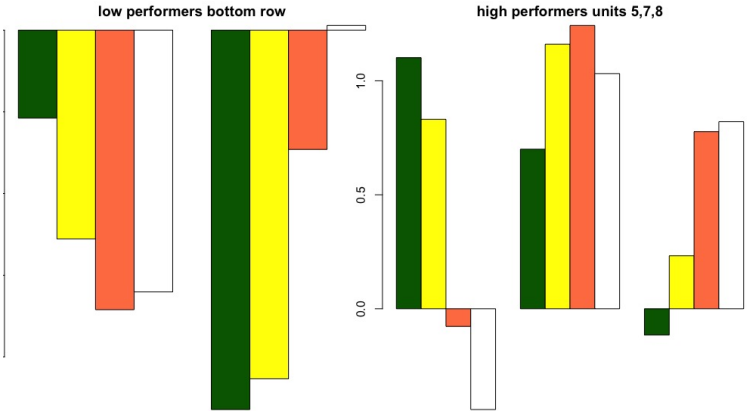


Figure 11: SOM 2-4 layout, codes for low and high performers' units

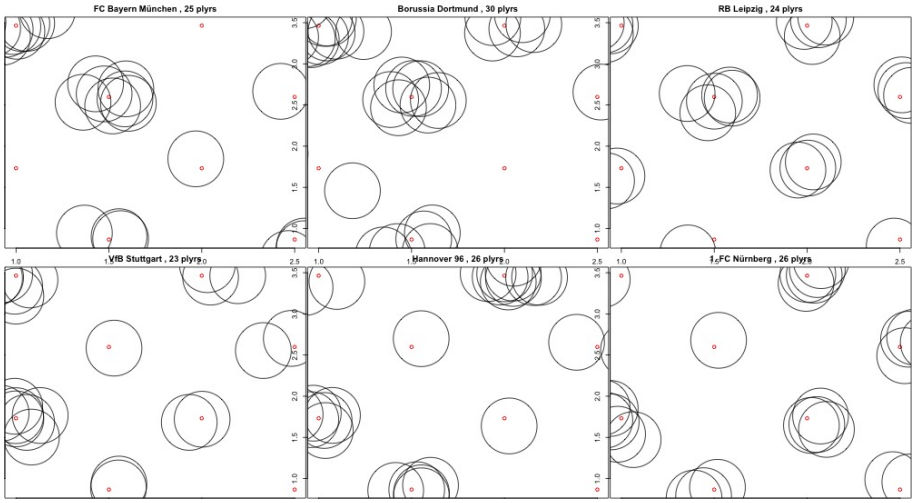
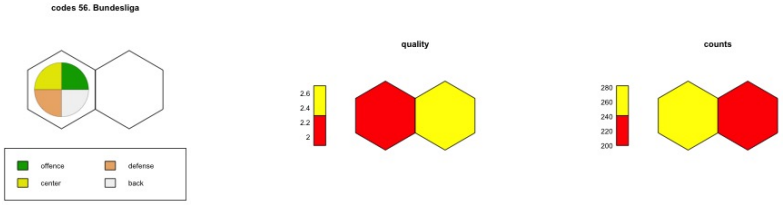
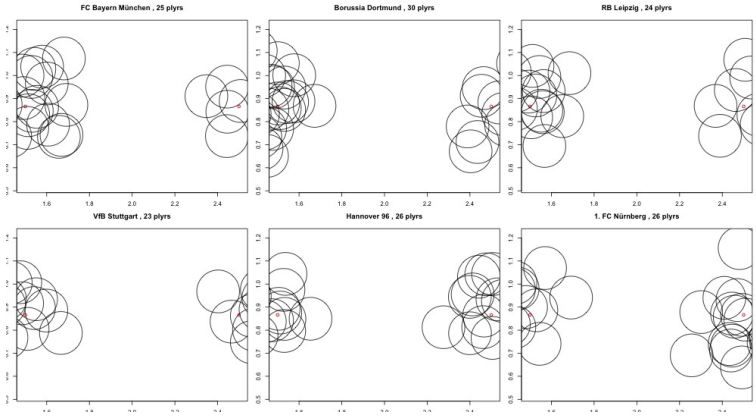


Figure 12: SOM 2-4 layout, clubs

5.3 SOM, extreme top or flop 2x1 layout





As for this 2-1 layout we do not really need the 2 dimensional som grid, we present a figure of small multiples of barcharts (titles missing, colour argh, arrangement should be more parallel to the figure above, all axes should be the same for comparability.).