

# Clustering of forwards based on game performance in the elite European football leagues

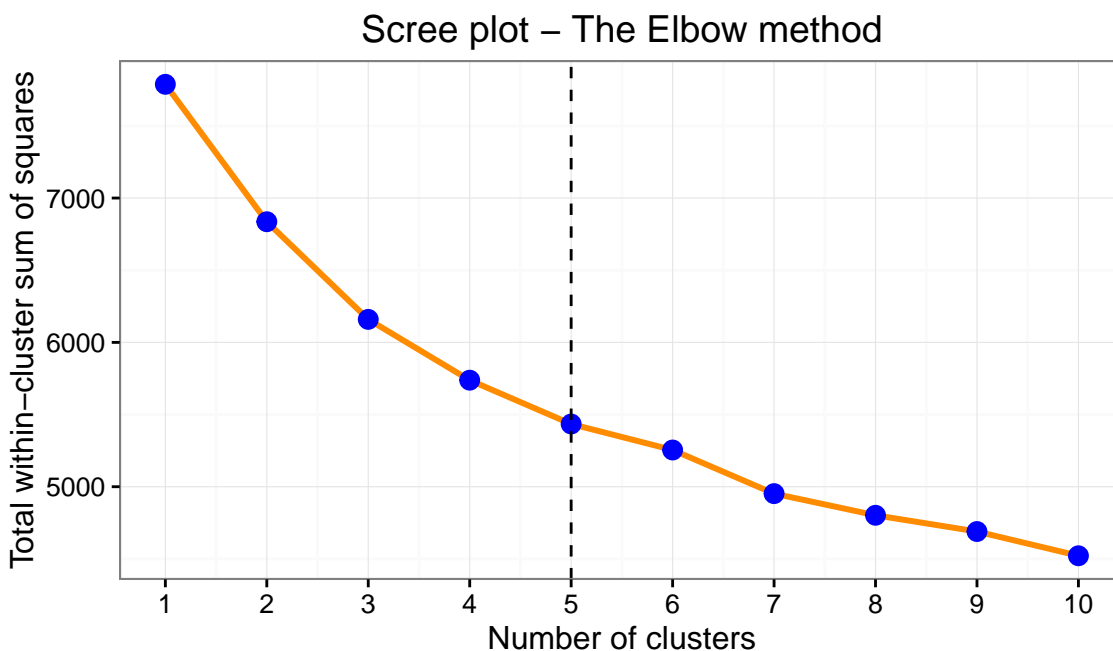
*Gustav Sternelöv*

*28 april 2016*

## Results

### Number of clusters

The  $k$  in the k-means algorithm is decided by looking at how the total within cluster sum of squares decreases as the number of clusters increases.



No evident elbow can be seen in the plot but that the improvement lowers as the number of clusters increases is clear. A  $k$  equal to either four or five is interpreted as being reasonable as the difference between five and six clusters is quite small. The difference between four and five clusters on the other hand is slightly clearer so a  $k$  is set to 5.

### The clusters

The size of the respective clusters are relatively even except for cluster two which only has nine players.

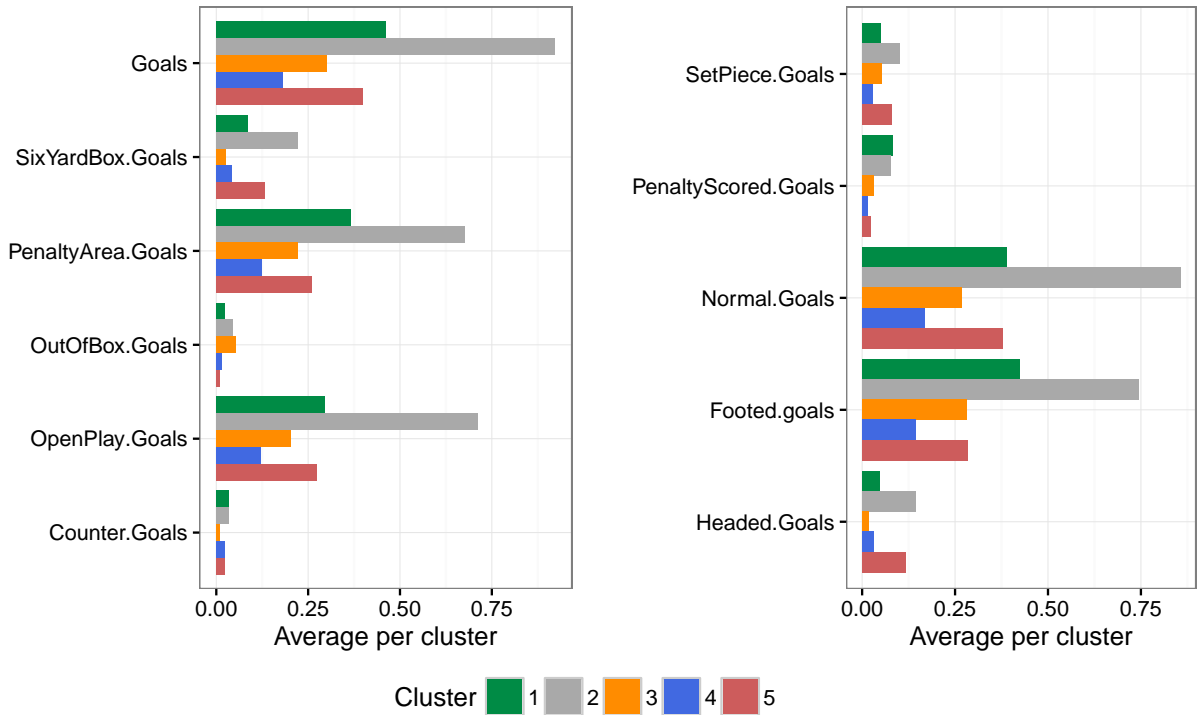
```
##
##  1  2  3  4  5
## 49  9 28 51 41
```

The five different clusters are compared by one category at the time. First, how the groups differ for the goal variables is presented and then they are compared given the variables in the categories shots, key passes and other.

## Goals

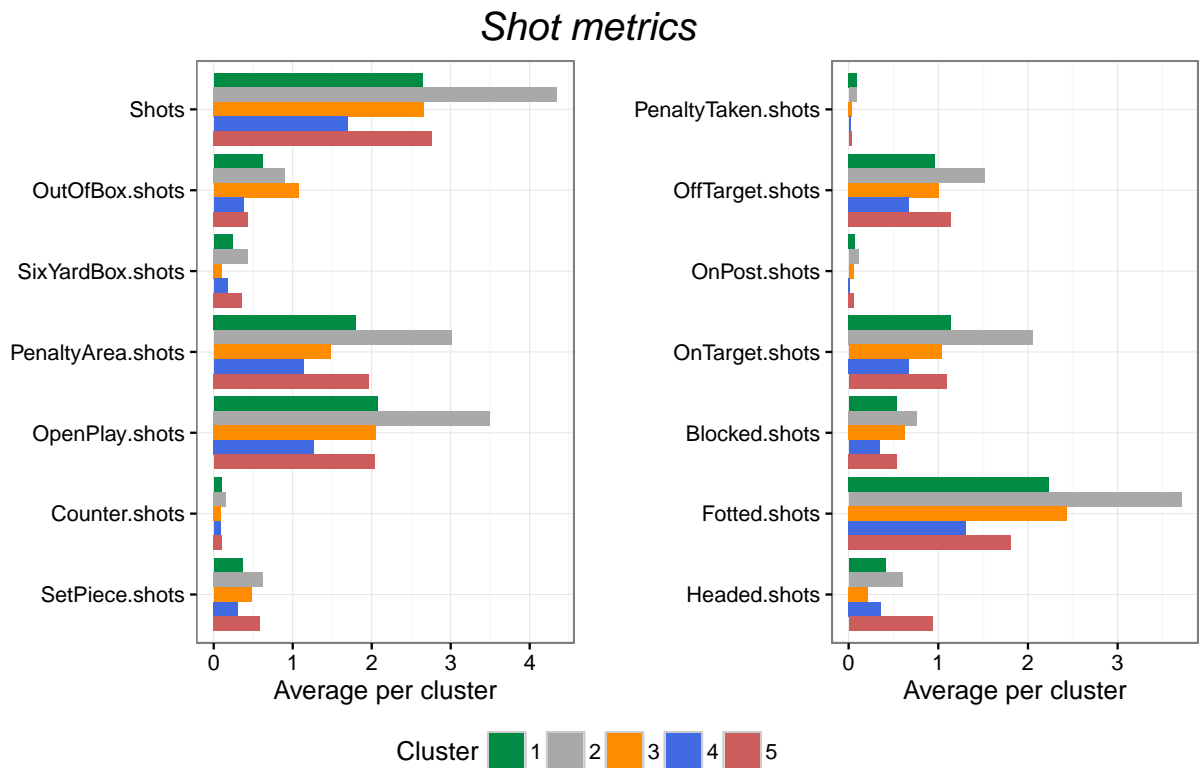
The average contribution from the players in the respective clusters is compared with the visualizations below. For each shot metric is the average number computed and represented with a bar in the graph.

### Goal metrics



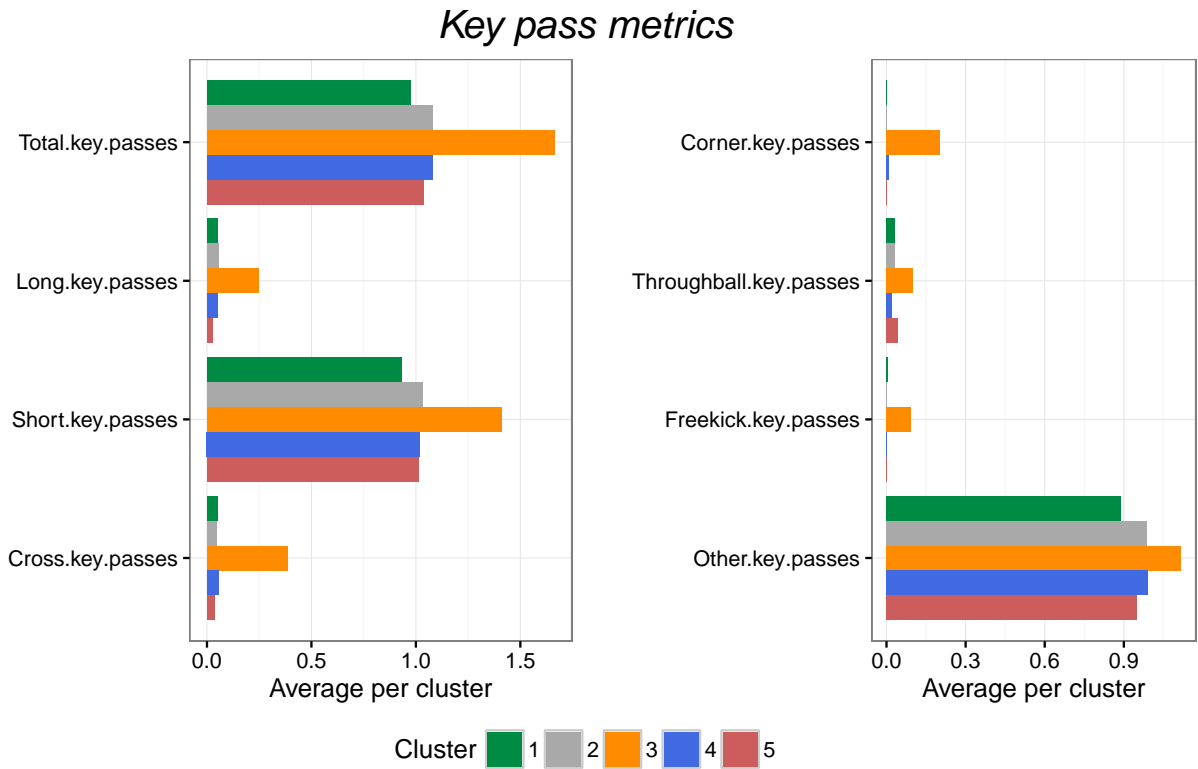
Cluster number 2, the grey one, includes the players that scores the most goals per 90 minutes and the average number is just below one. Cluster 1, the green, comes next and thereafter follows cluster 5, the red, and cluster 3, the orange. The lowest number of goals per 90 minutes has the players in cluster 4, the blue one. For most of the metrics can the same order and relative difference between the clusters be observed. However, deviations from that pattern can be seen for a few metrics. Players that are clustered into the orange cluster very seldom scores in the six-yard box. Instead, they score relatively often with shots from outside the box. Another exception can be seen for the players in the red cluster since a notably big part of their goals are headed.

## Shots



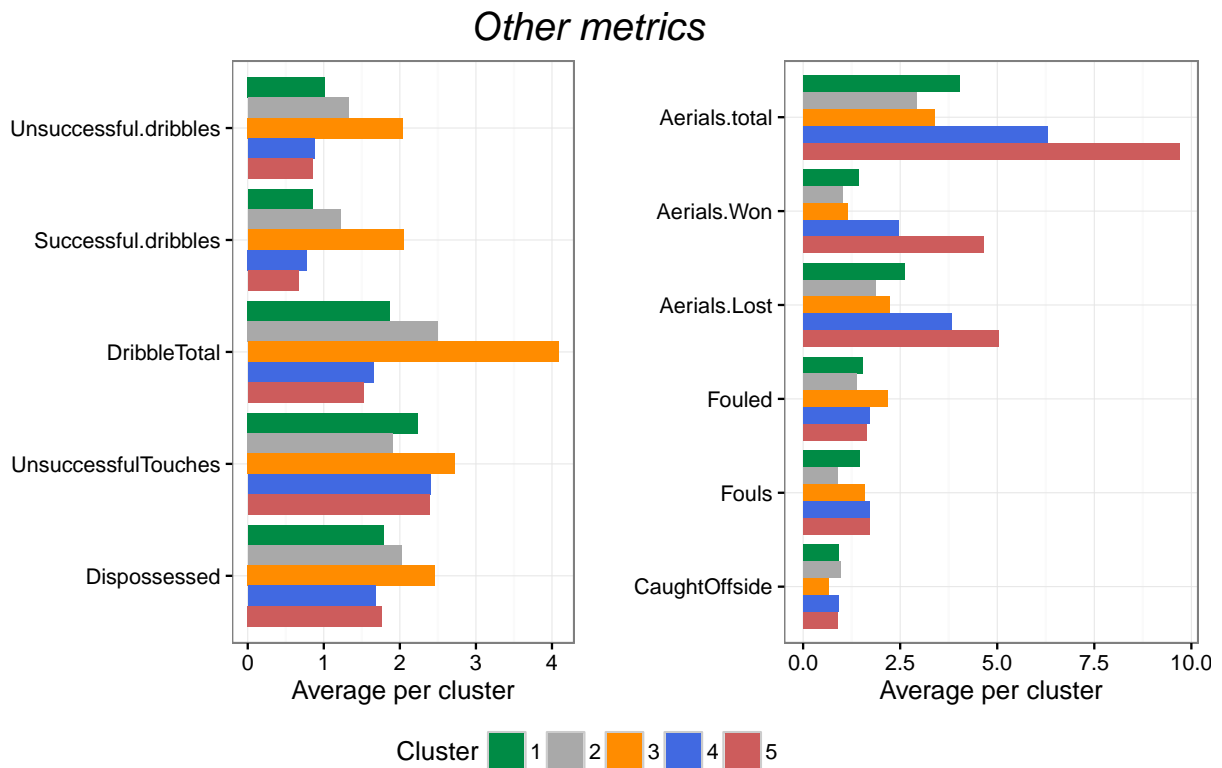
The forwards in the grey cluster has the highest amount of shots per 90 minutes played. Thereafter follows again the green, orange and red cluster and it is really close between the averages of these clusters. The blue cluster has the lowest average number of shots. The metrics for which this pattern cannot be seen is Out-of box shots and headed shots. In average does the players in the orange cluster take a lot of shots from outside the box and the players in the red cluster has quite many headed shots per 90 minutes. The general pattern for the green and blue cluster is similar to the pattern for the grey cluster, with the exception that their average number is lower.

## Key passes



The average number of key passes is by margin highest for the orange cluster. Between the other clusters it is very even. None of the averages for the other metrics give any especially notable information apart from that the highest average is found for the orange cluster.

## Other



## Lewandowski's replacement

Lewandowski is in cluster two, the cluster with only nine players in total. Exclusive group of players who on average scores around one goal per game and takes 4-5 shots per game. A bit more spread out for the key passes. The big common factor is the number of chances these players creates for themselves and how well they converts the chances into goals.

##	Players	Goals90	Shots90	KeyPasses90
## 1	Harry Kane	0.7	4.2	1.2
## 2	Sergio Agüero	0.9	4.5	1.0
## 3	Daniel Sturridge	0.7	4.2	0.6
## 4	Robert Lewandowski	0.8	4.8	0.7
## 5	Pierre-Emerick Aubameyang	0.9	4.2	1.0
## 6	Gonzalo Higuaín	1.1	5.5	1.5
## 7	Luis Suárez	1.1	3.9	1.6
## 8	Karim Benzema	1.1	4.4	1.9
## 9	Imanol Agirretxe	1.0	3.4	0.2

## Discussion