



FOOTBALL BETTING TIPS

CROON NICOLA
MOLTINI SIBILLA

AGENDA

1. RESEARCH QUESTION

2. DATA VISUALIZATION AND EXPLORATION

3. PCA

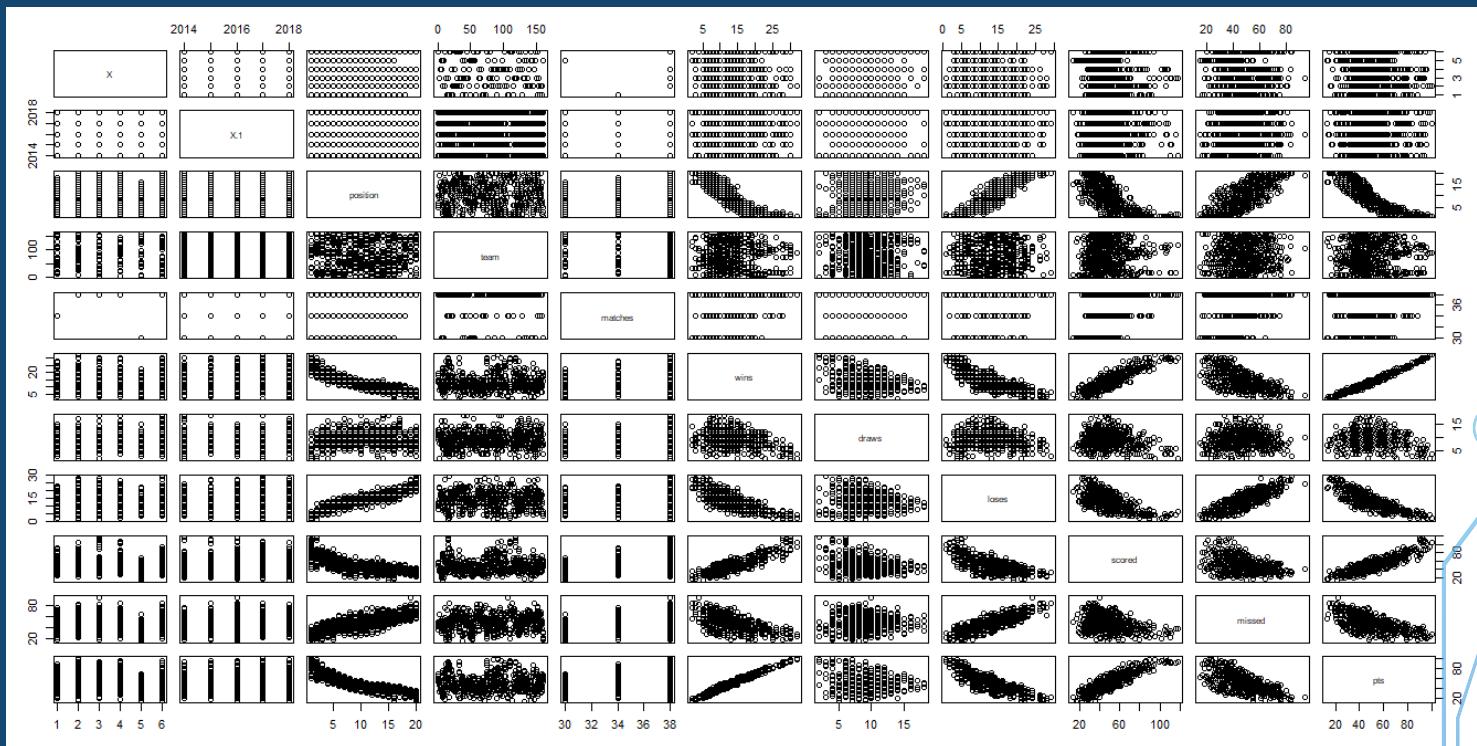
4. CLUSTERING

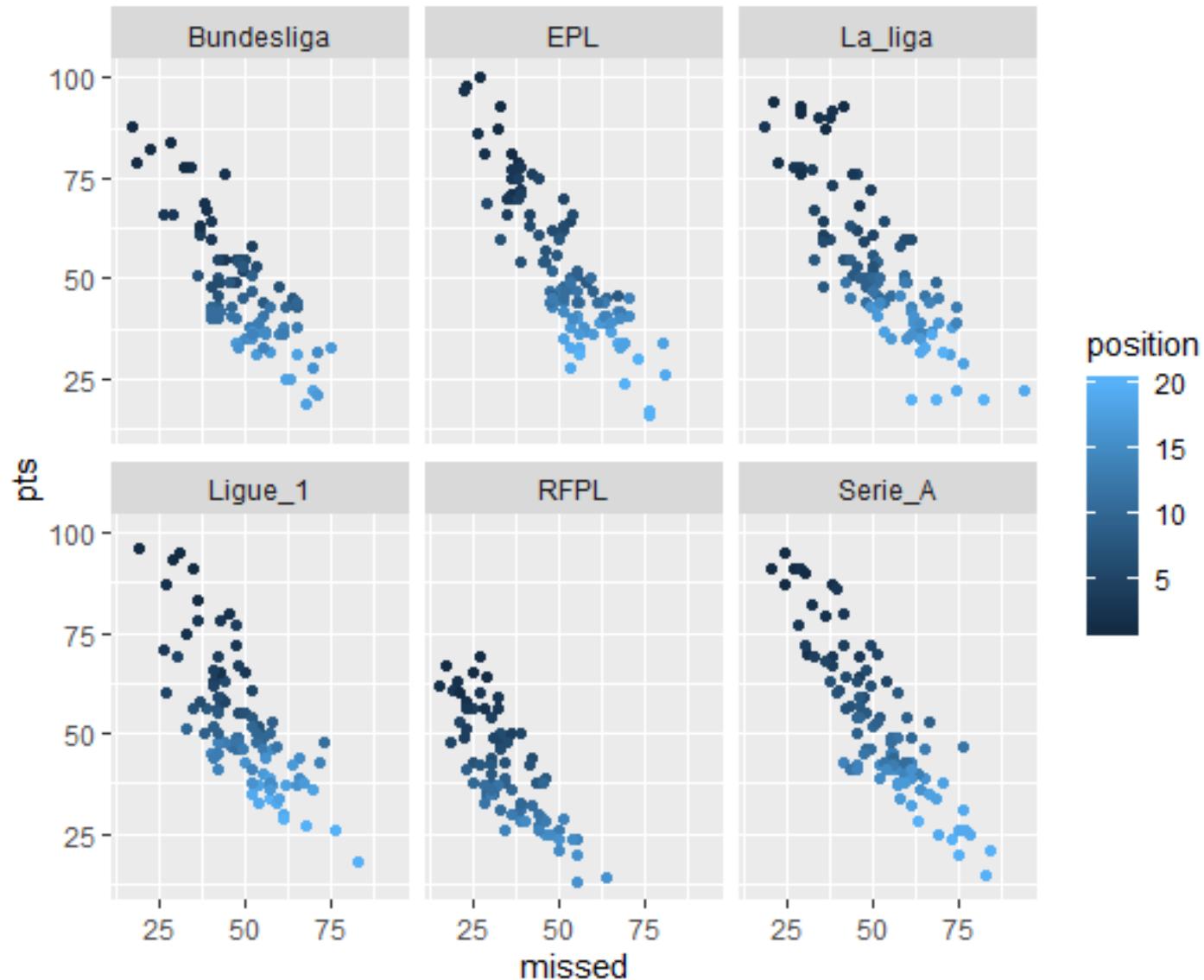
5. CONCLUSION AND POSSIBLE DEVELOPMENTS

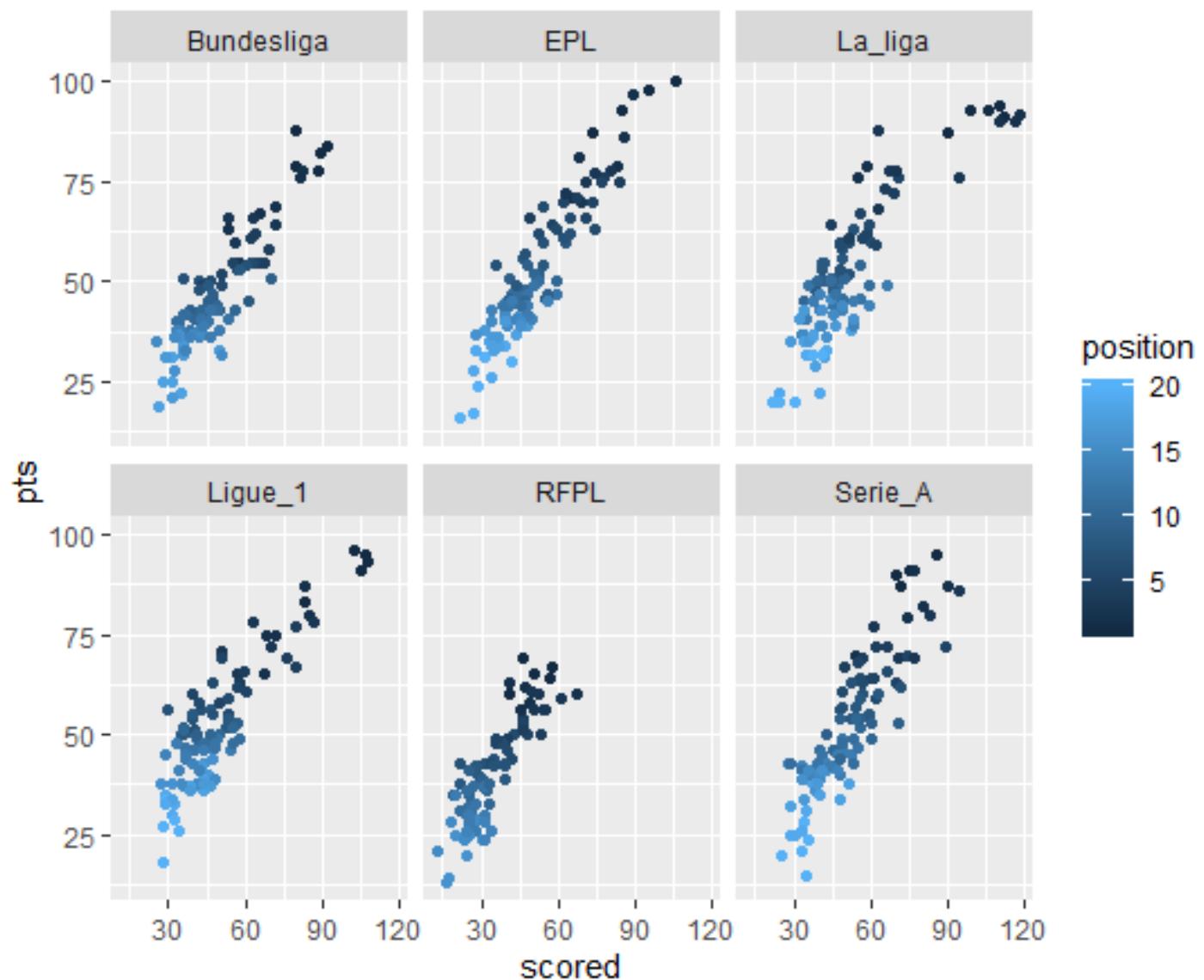
THE DATA

- **PARAMETERS:**
- **Position**
- **Matches played**
- **Missed goals**
- **Scored goals**
- **Teams**
- **Points**
- **Wins**
- **Draws**
- **Loses**

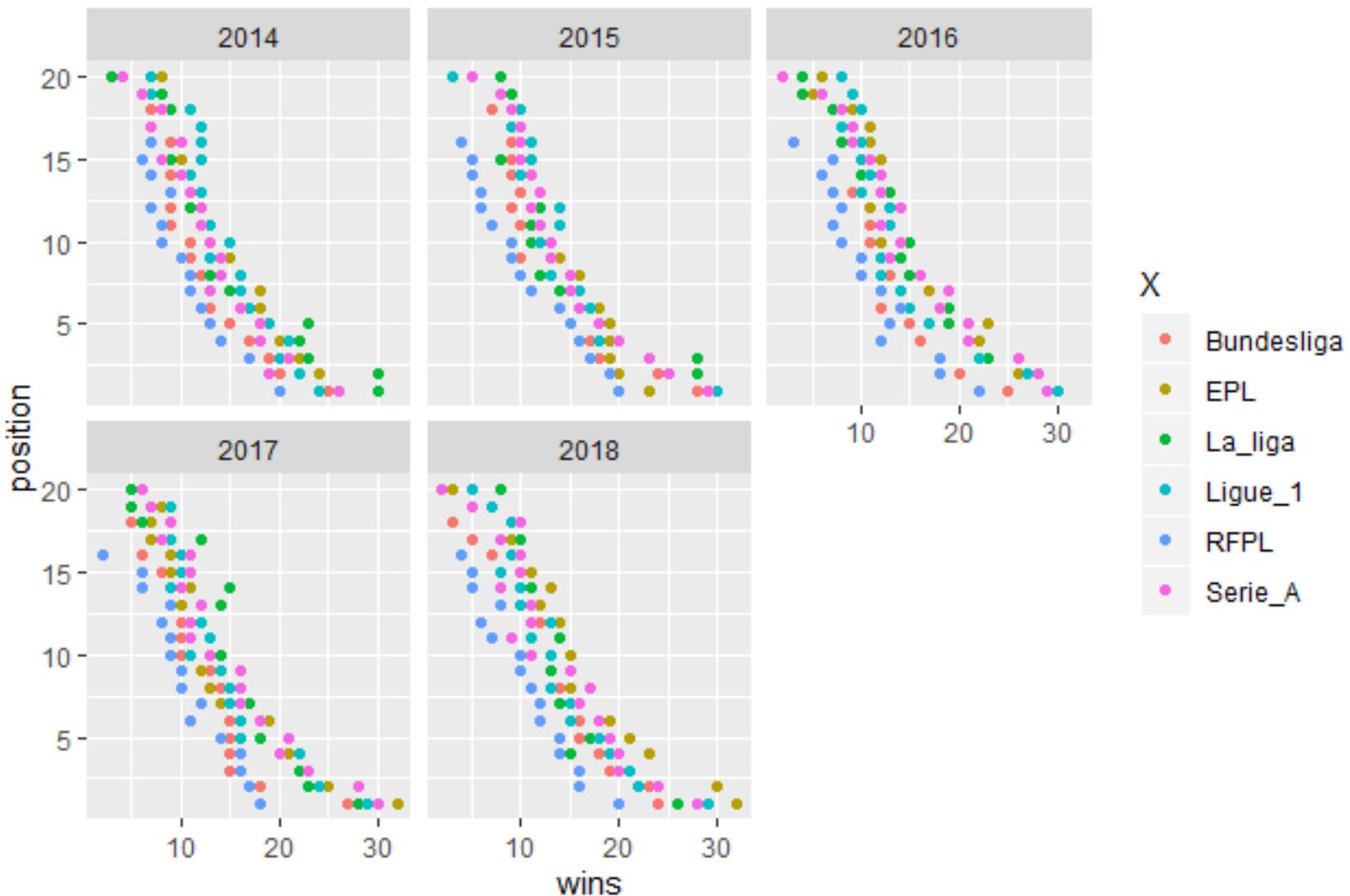
<https://www.kaggle.com/slehkyi/extended-football-stats-for-european-leagues-xg>



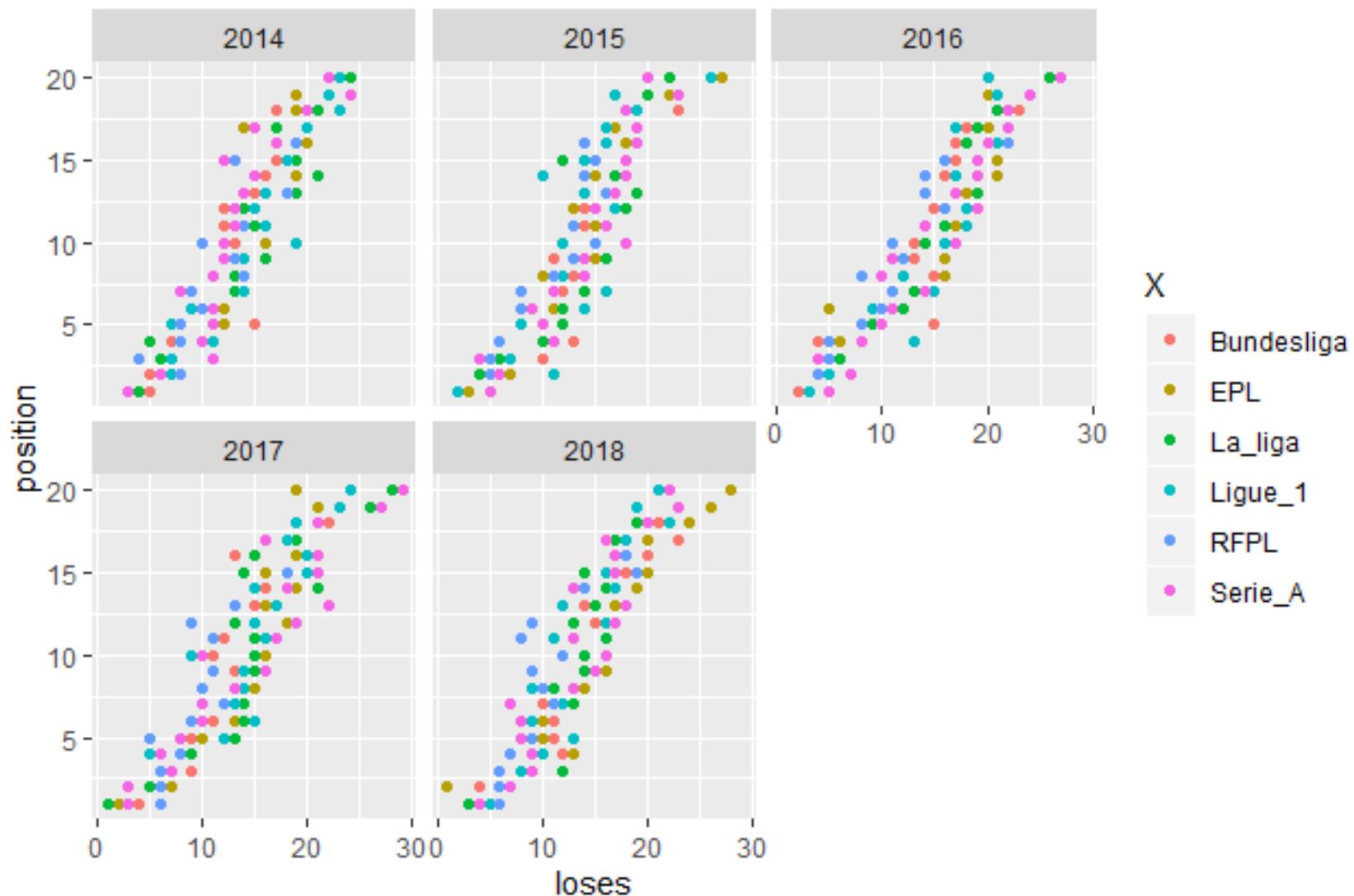




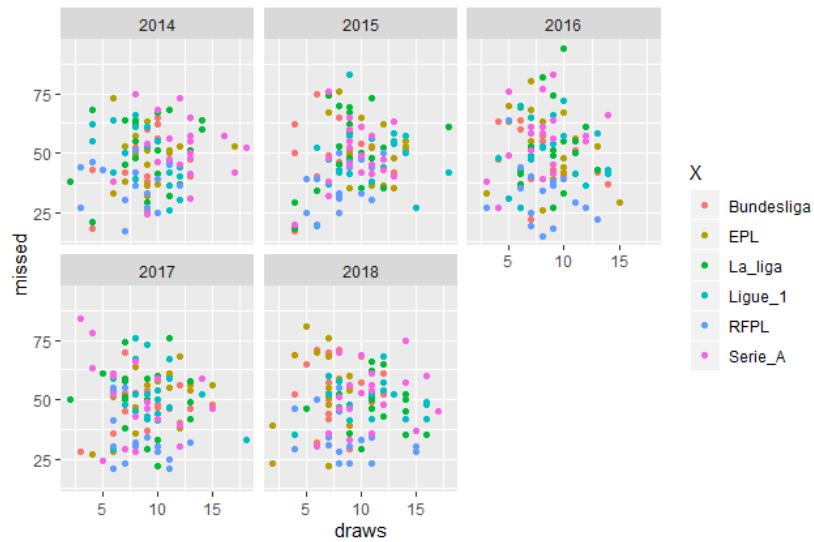
wins and position



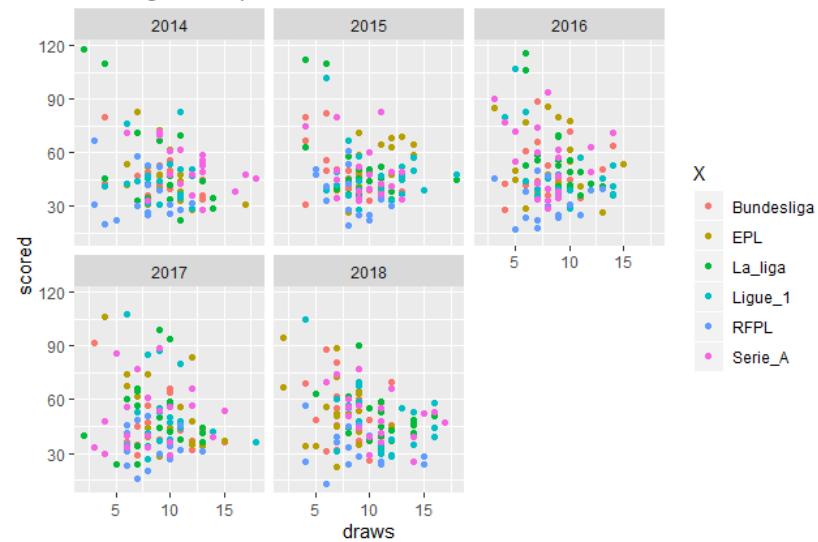
lost matches and consequent position



do the goals help?

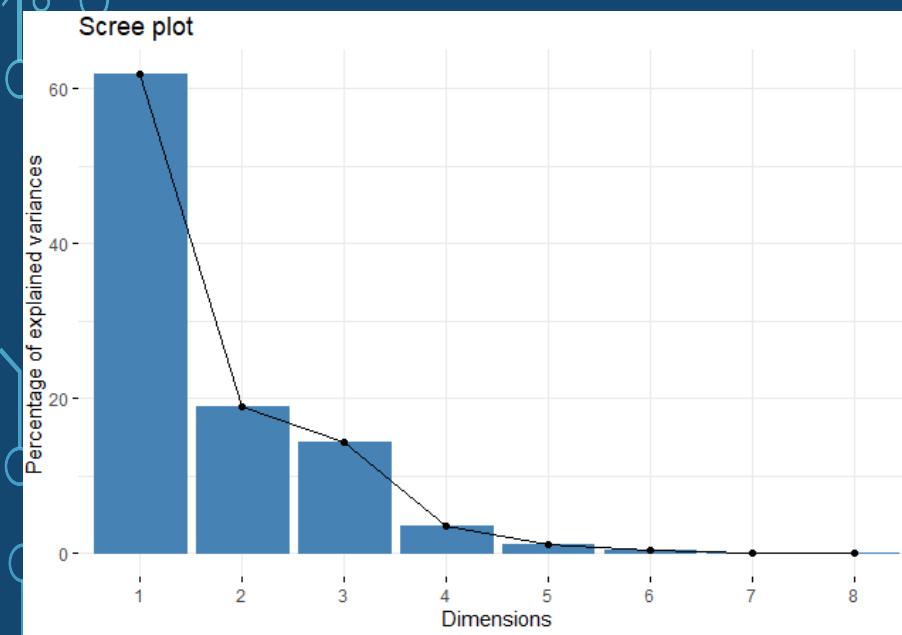


do the goals help?

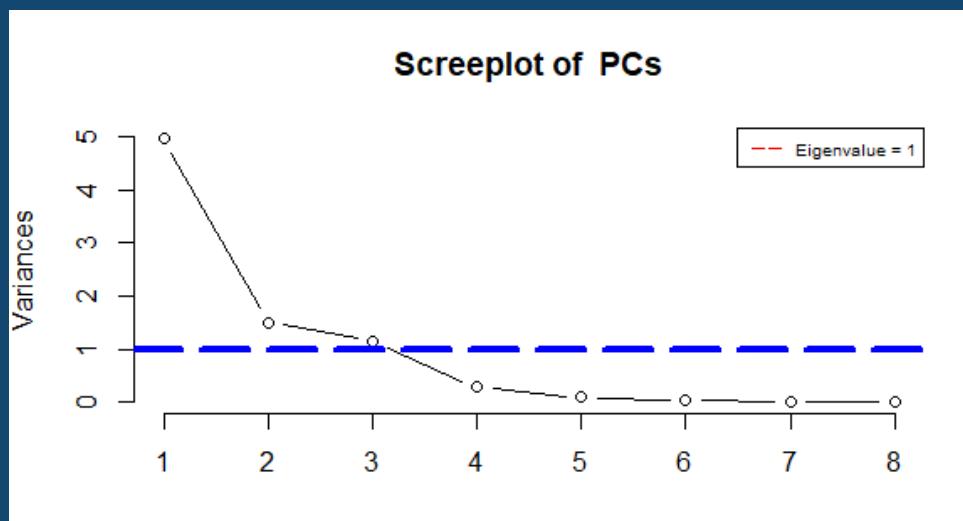


PRINCIPAL COMPONENT ANALYSIS

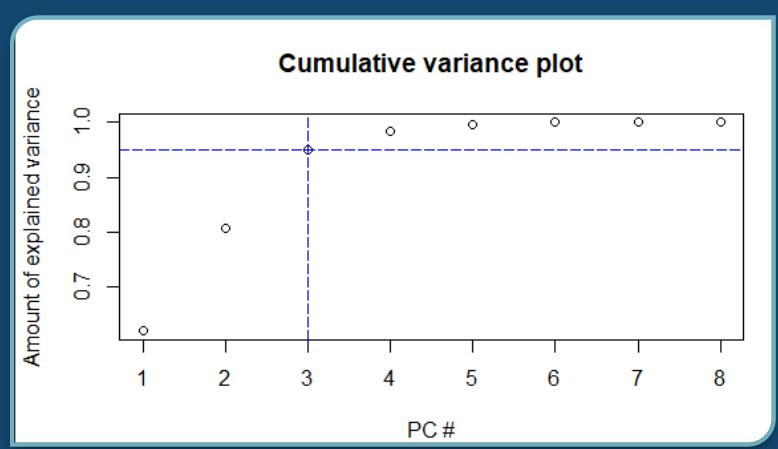
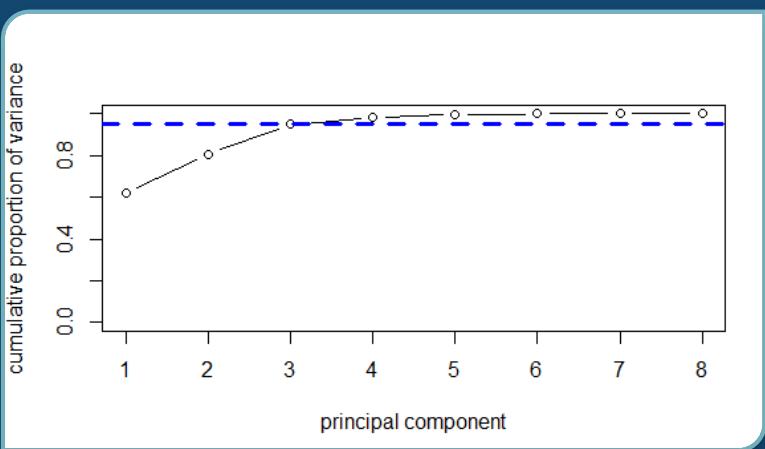
Scree plot



Screeplot of PCs

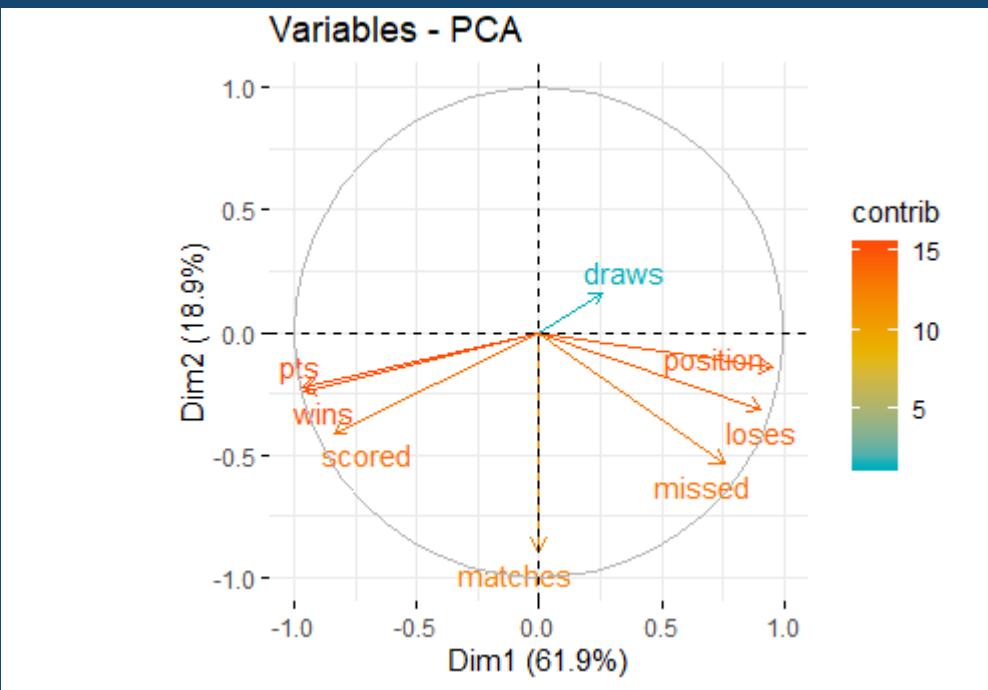


CUMULATIVE VARIANCE



```
Importance of components:  
PC1 PC2 PC3 PC4 PC5 PC6 PC7  
Standard deviation 2.2248 1.2280 1.0678 0.52951 0.29870 0.18072 3.878e-16  
Proportion of Variance 0.6187 0.1885 0.1425 0.03505 0.01115 0.00408 0.000e+00  
Cumulative Proportion 0.6187 0.8072 0.9497 0.98476 0.99592 1.00000 1.000e+00  
PC8  
Standard deviation 1.741e-16  
Proportion of Variance 0.000e+00  
Cumulative Proportion 1.000e+00
```

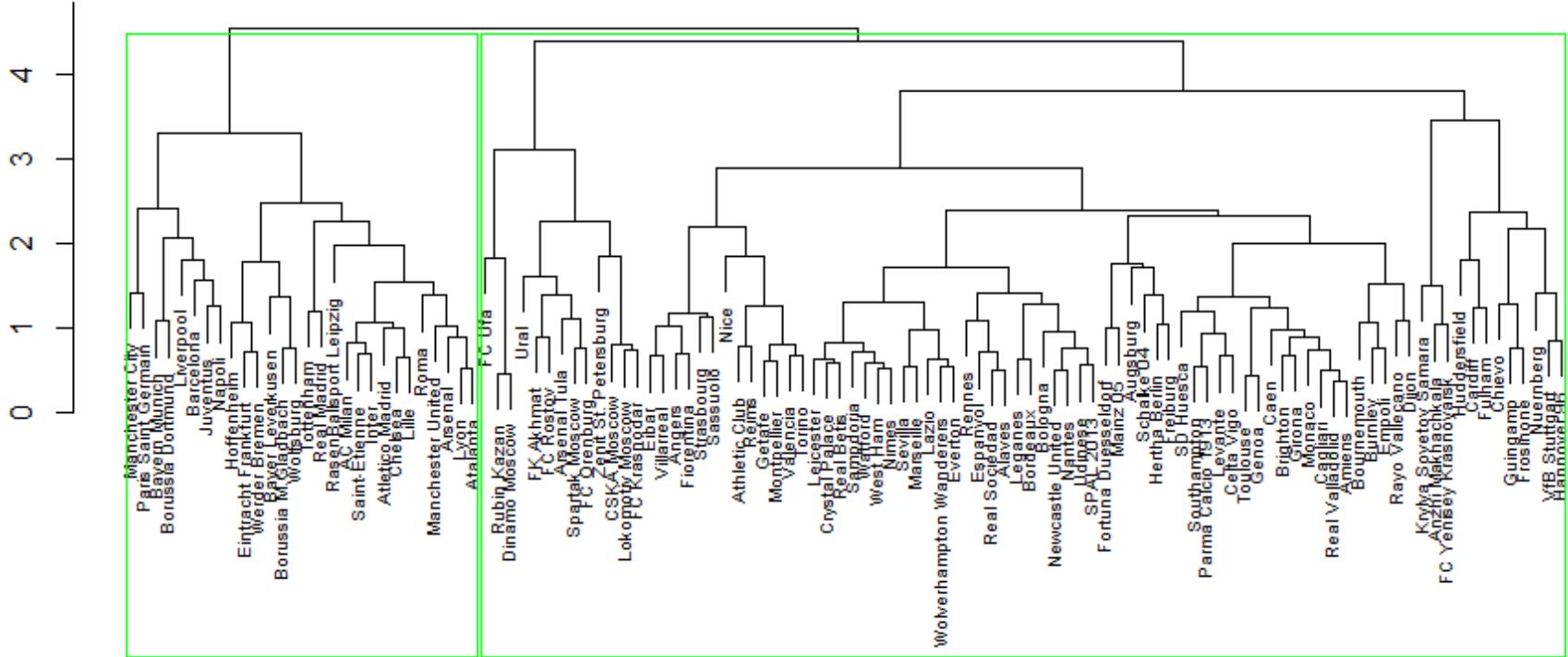
BI PLOT



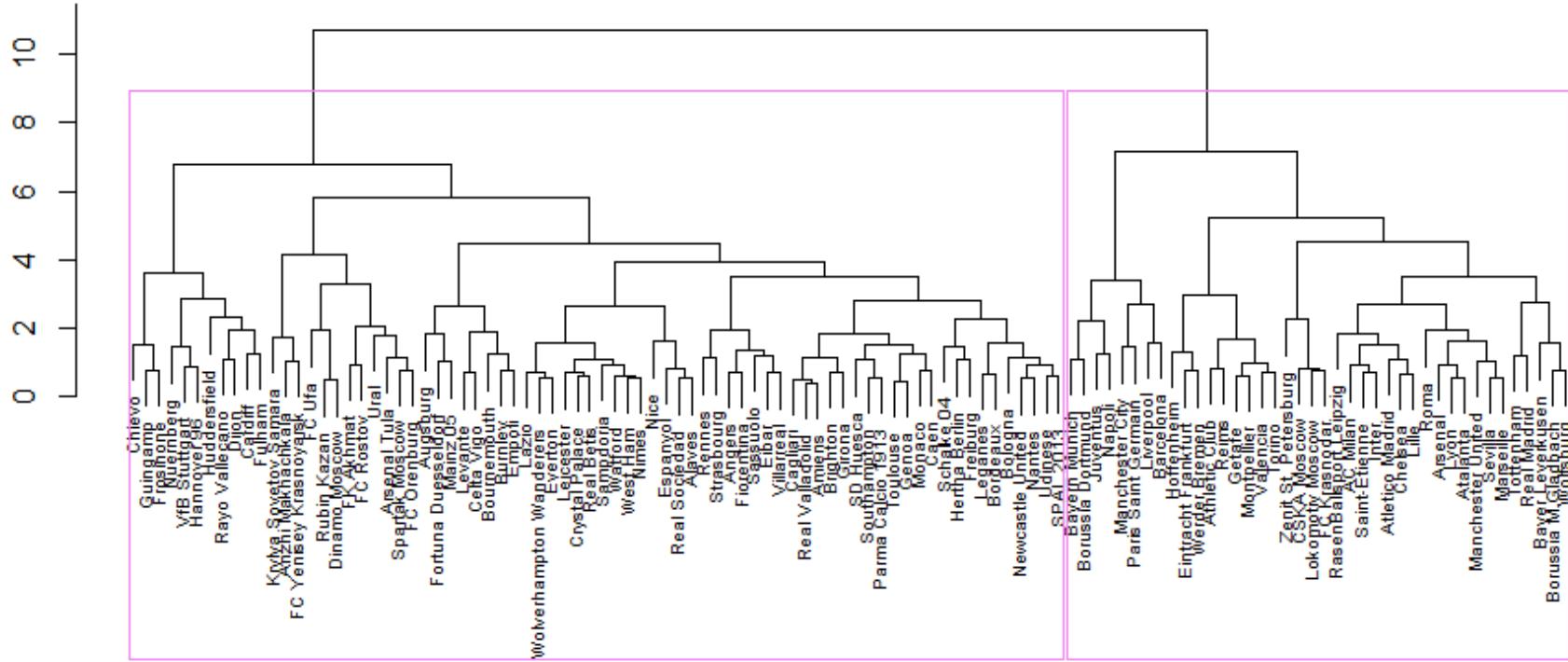
HIERARCHICAL CLUSTERING

- SEEK FOR A PARTITION OF THE DATA SUCH THAT THE OBSERVATION WITHIN EACH GROUP ARE SIMILAR TO EACH OTHER.
- THIS METHOD CLUSTERS THE DATA SEQUENTIALLY STARTING WITH ALL THE OBSERVATIONS, THAT ARE THE OBSERVATIONS OF THE YEAR 2018
- EUCLIDEAN DISTANCE IS THE DISSIMILARITY MEASURE THAT WE WILL CONSIDER
- LABELS: TEAMS

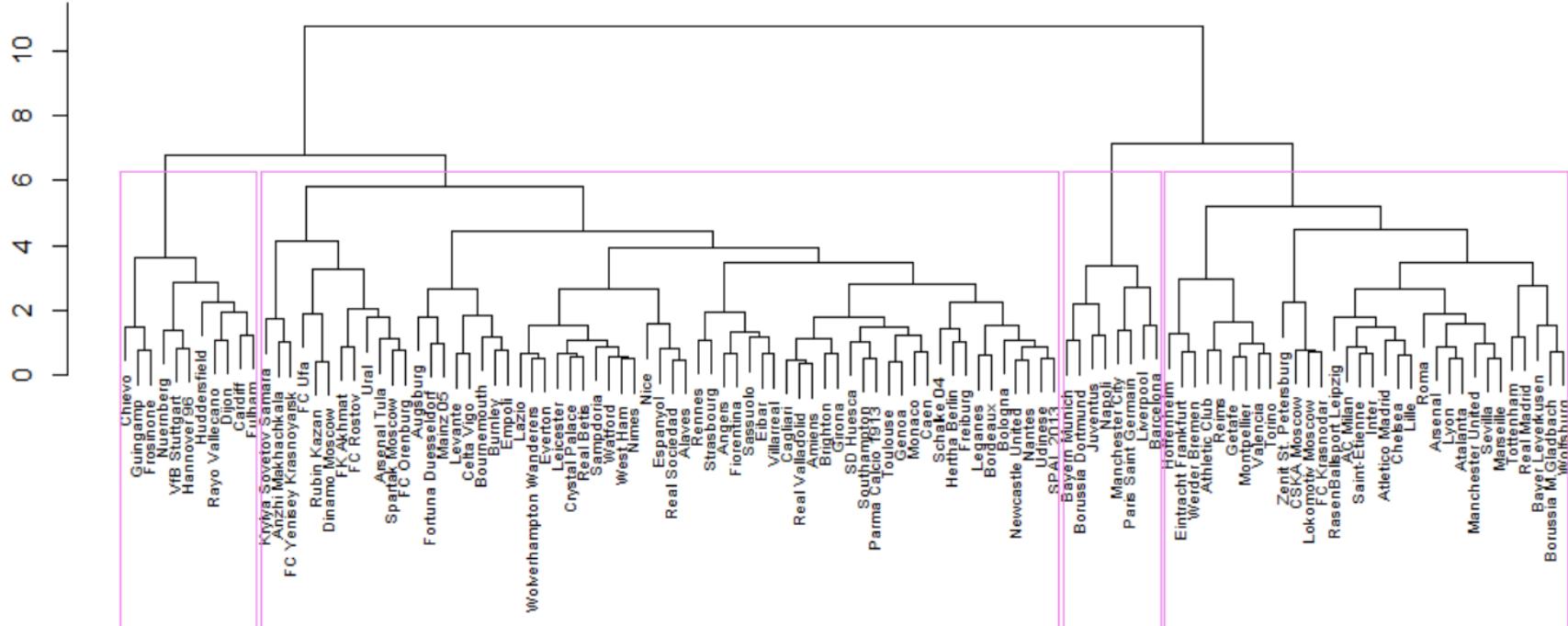
HIERARCHICAL CLUSTERING –AVERAGE LINKAGE



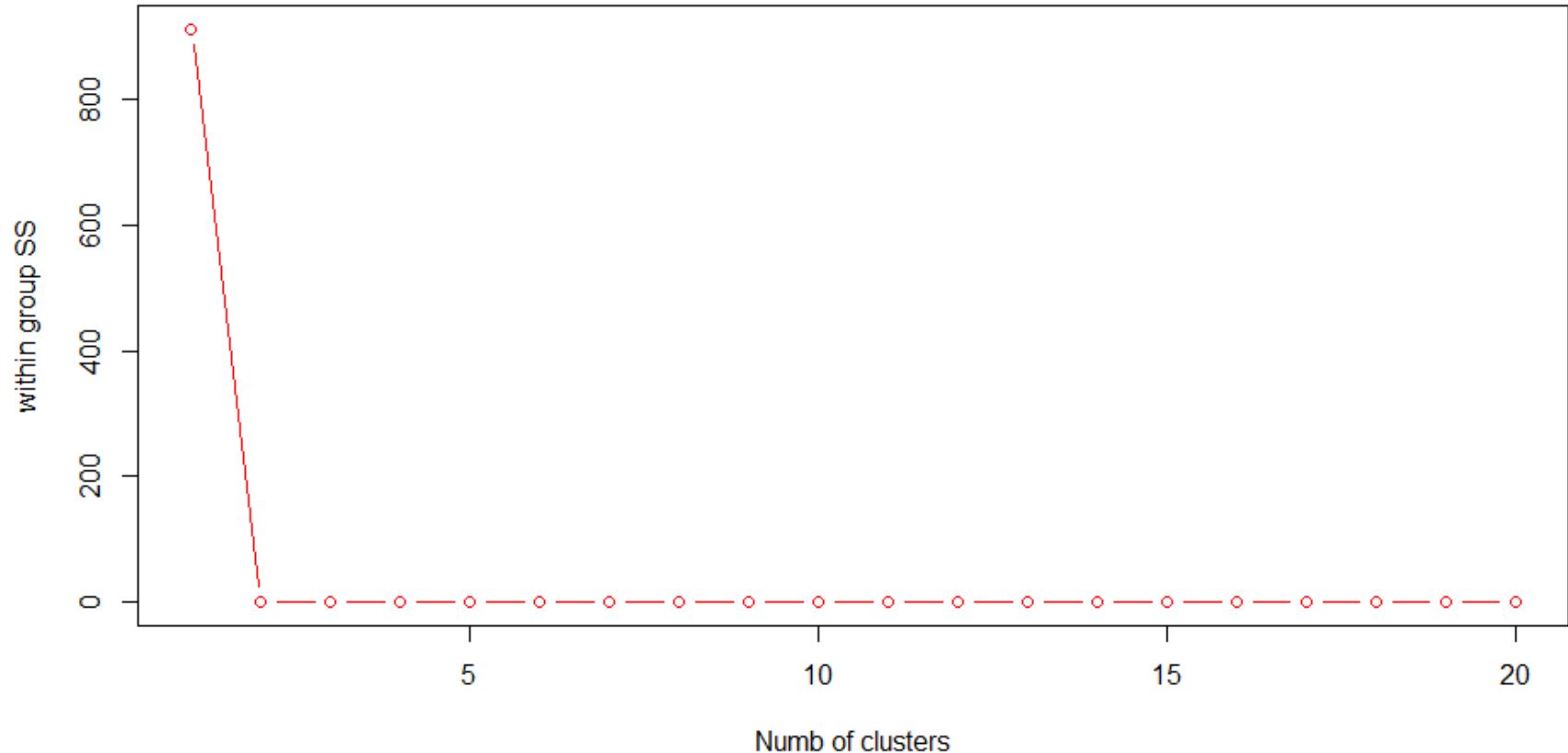
HIERARCHICAL CLUSTERING – COMPLETE LINKAGE



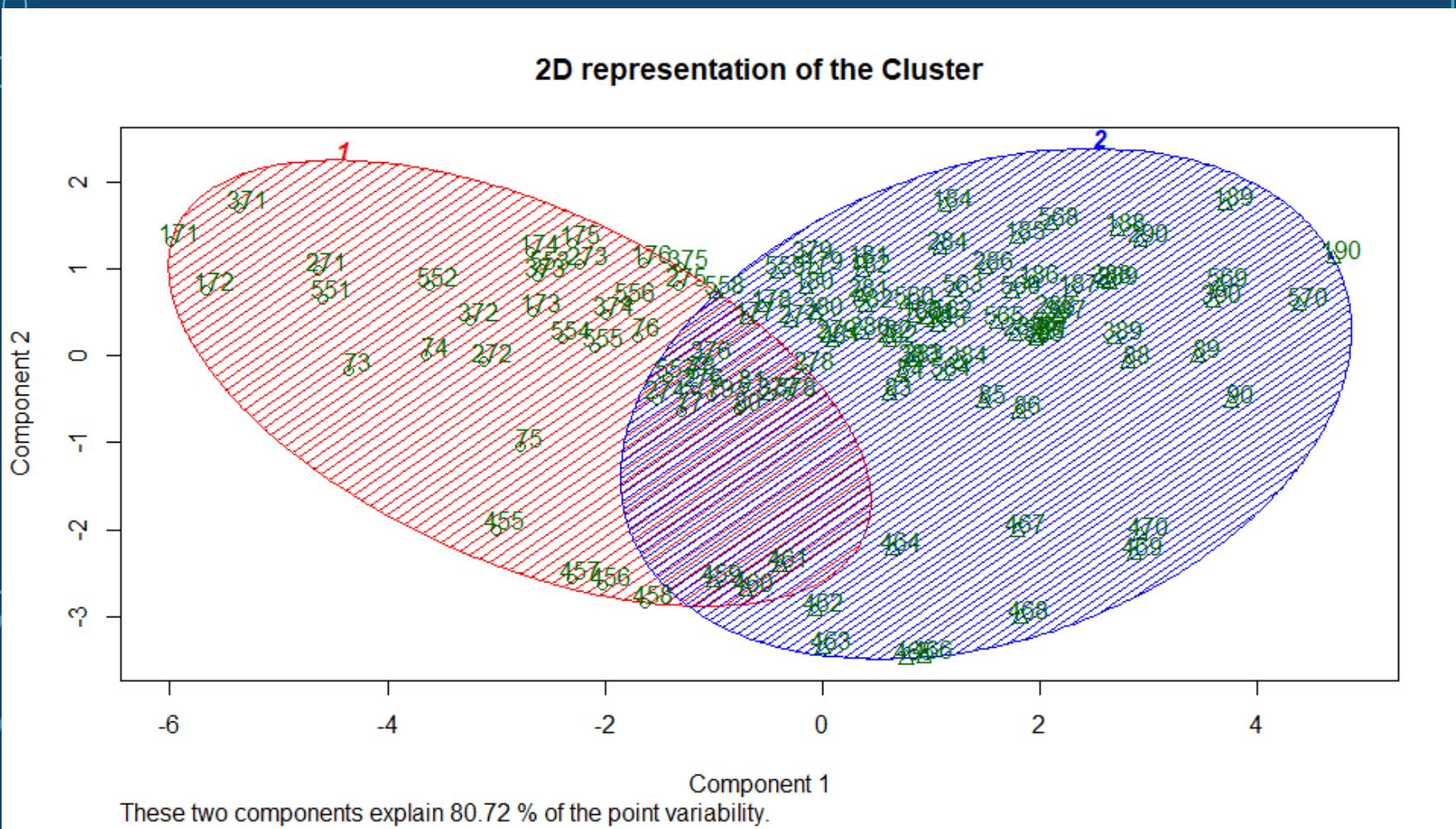
HIERARCHICAL CLUSTERING – COMPLETE LINKAGE



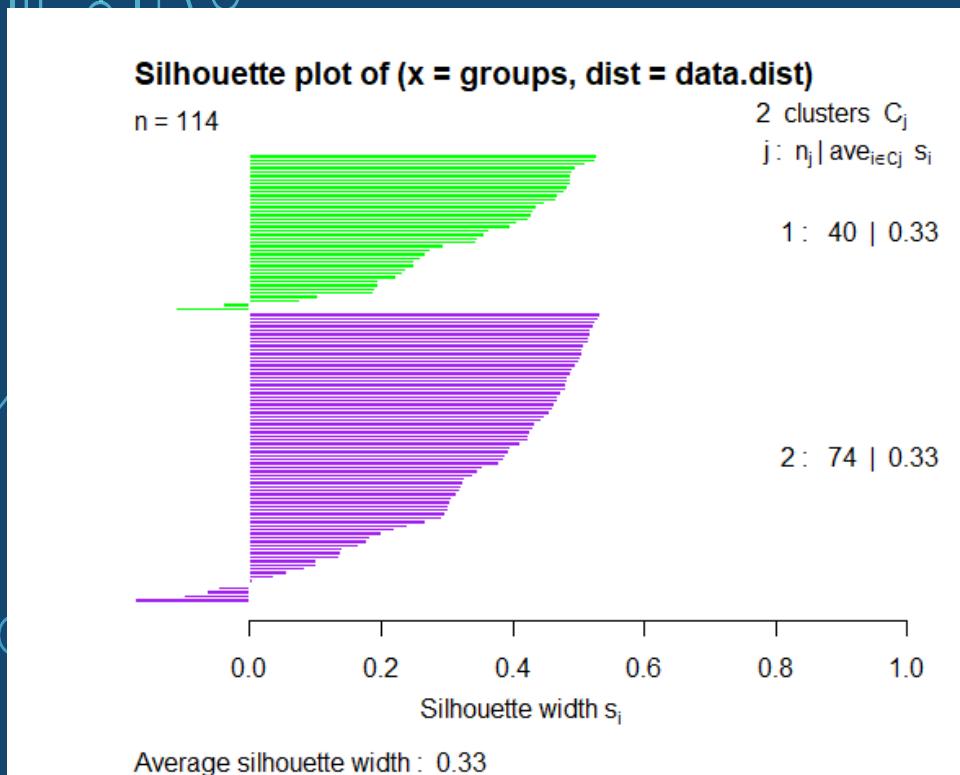
SCREE PLOT – COMPLETE LINKAGE



CUSPLOT 2D REPRESENTATION



SILHOUETTE COEFFICIENTS

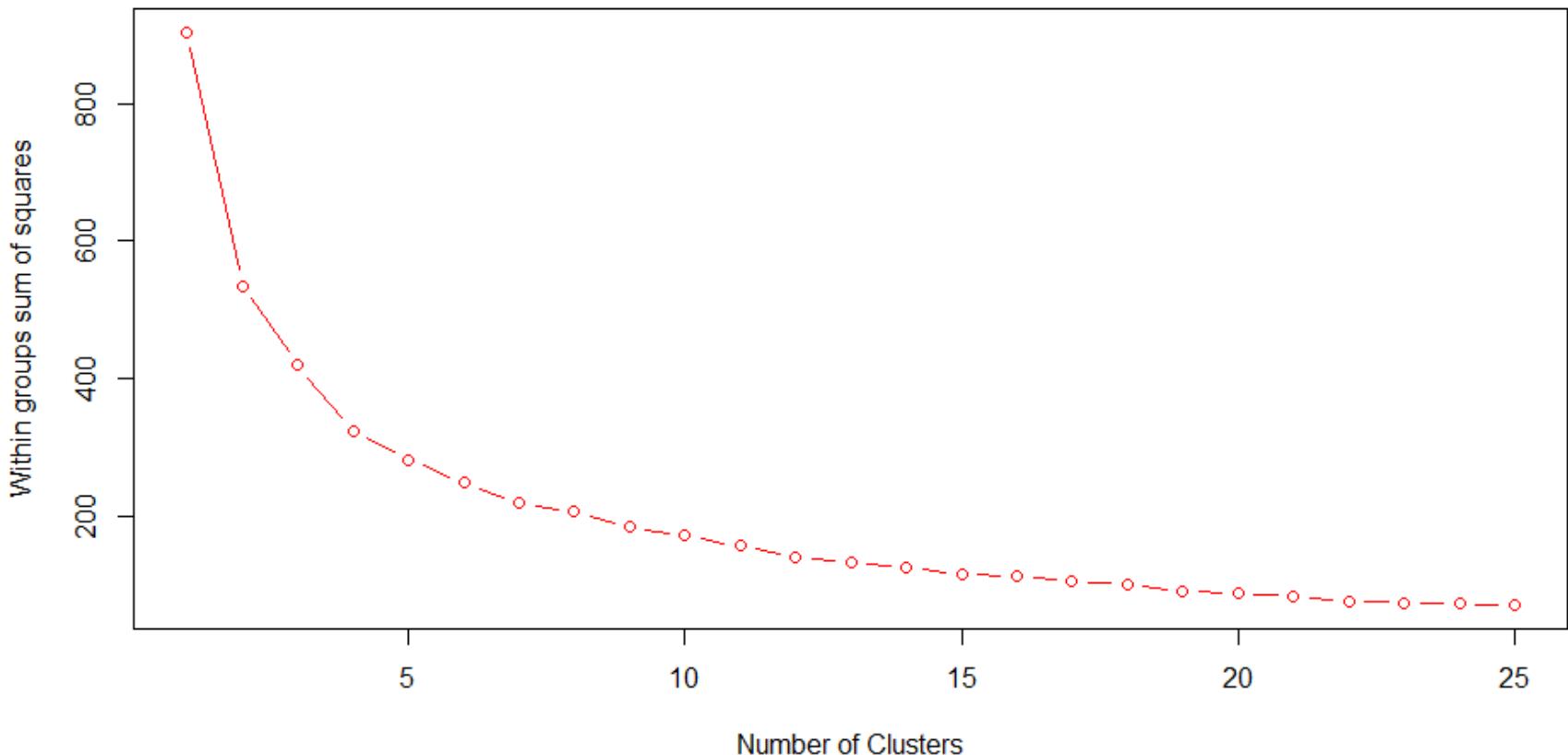


- «S» IS THE SILHOUETTE COEFFICIENT FOR EACH OBSERVATION.
- POSITIVE S MEANS THAT THE OBSERVATION IS ON AVERAGE CLOSER TO THE POINTS IN ITS OWN CLUSTER
- NEGATIVE MEANS THAT THE OBS IS CLOSER TO THE OTHER CLUSTER
- WE OBTAINED GOOD RESULTS!
- AVERAGE VALUE 0.33

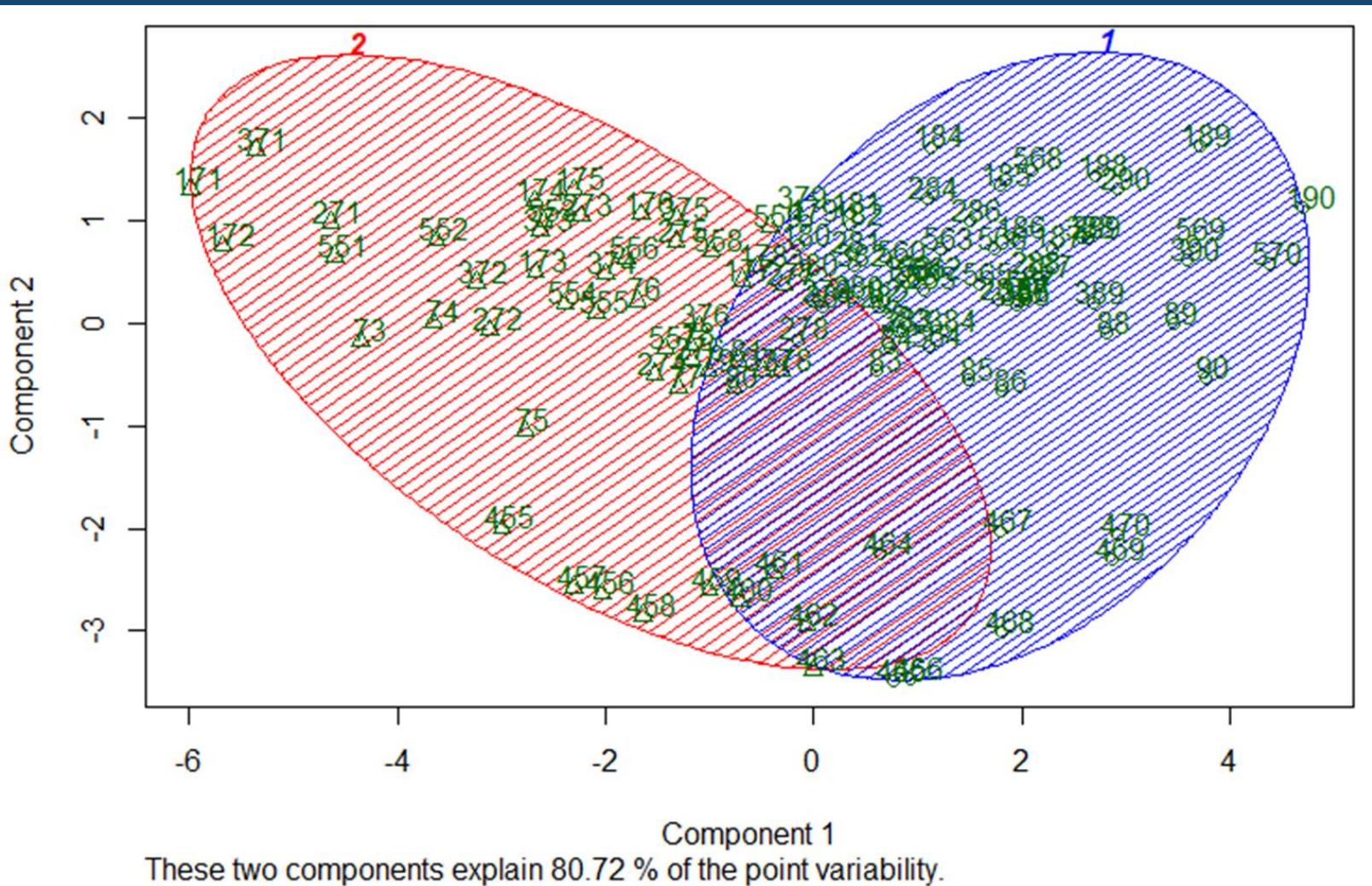
K – MEANS CLUSTERING

- THIS METHOD STARTS BY PICKING UP «K» INITIAL CLUSTER EXEMPLARS (MEANS)
 1. KEEPING MEANS FIXED WE ASSIGN EACH POINT TO THE CLOSEST MEAN
 2. KEEPING THE ASSIGNMENT FIXED MOVE EACH EXEMPLARS TO THE CENTRE OF ITS ASSIGNED CLUSTER
- THE INITIATION MATTER: WE STARTED WITH K=2 AND WE FOUND OUT THE WITHIN CLUSTER SUM OF SQUARES
 1. 310.4444
 2. 224.6392
- WE END UP WITH K=4
 1. 103.01647
 2. 98.75702
 3. 37.95619
 4. 86.50920

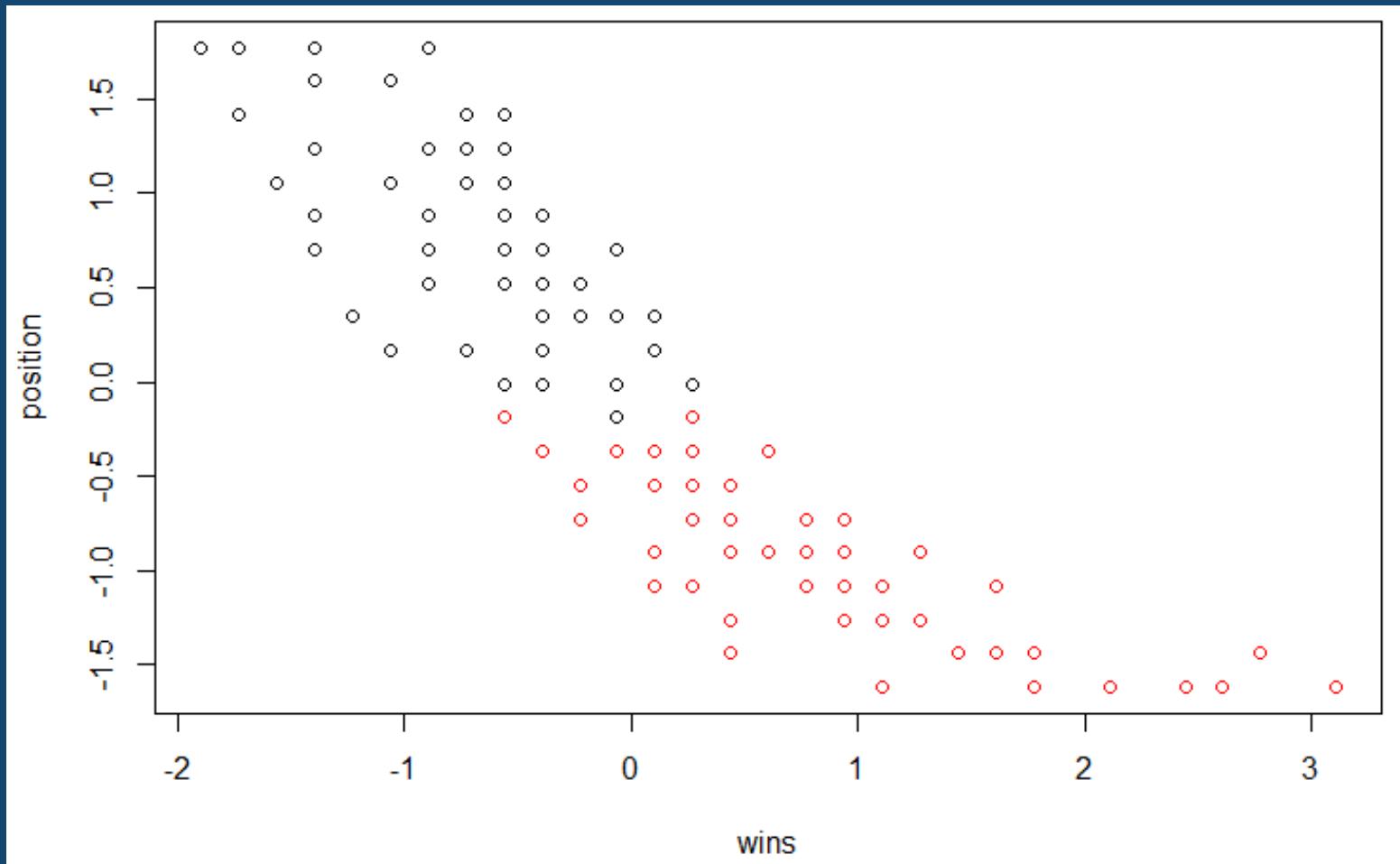
SCREE PLOT



CUSPLOT 2D REPRESENTATION WITH K=2

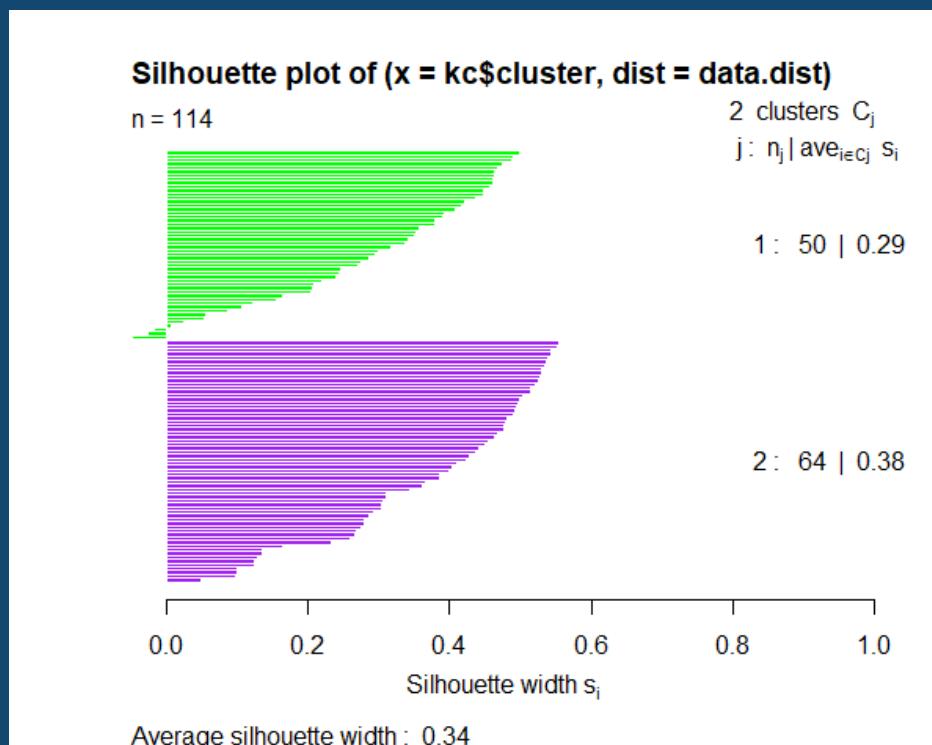


EXAMPLE OF DATA VISUALIZATION WITH K=2

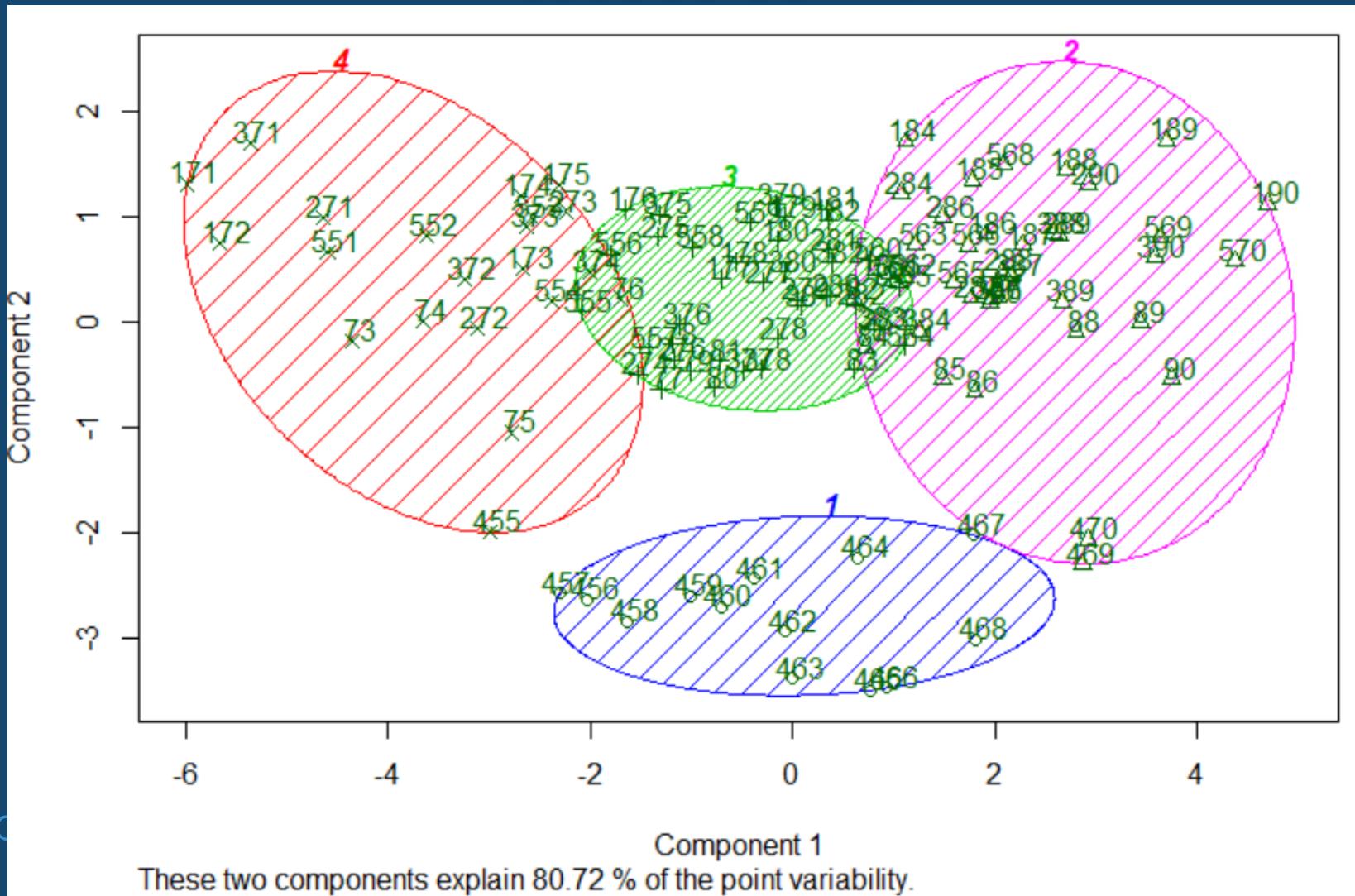


SILHOUETTE COEFFICIENTS FOR K=2

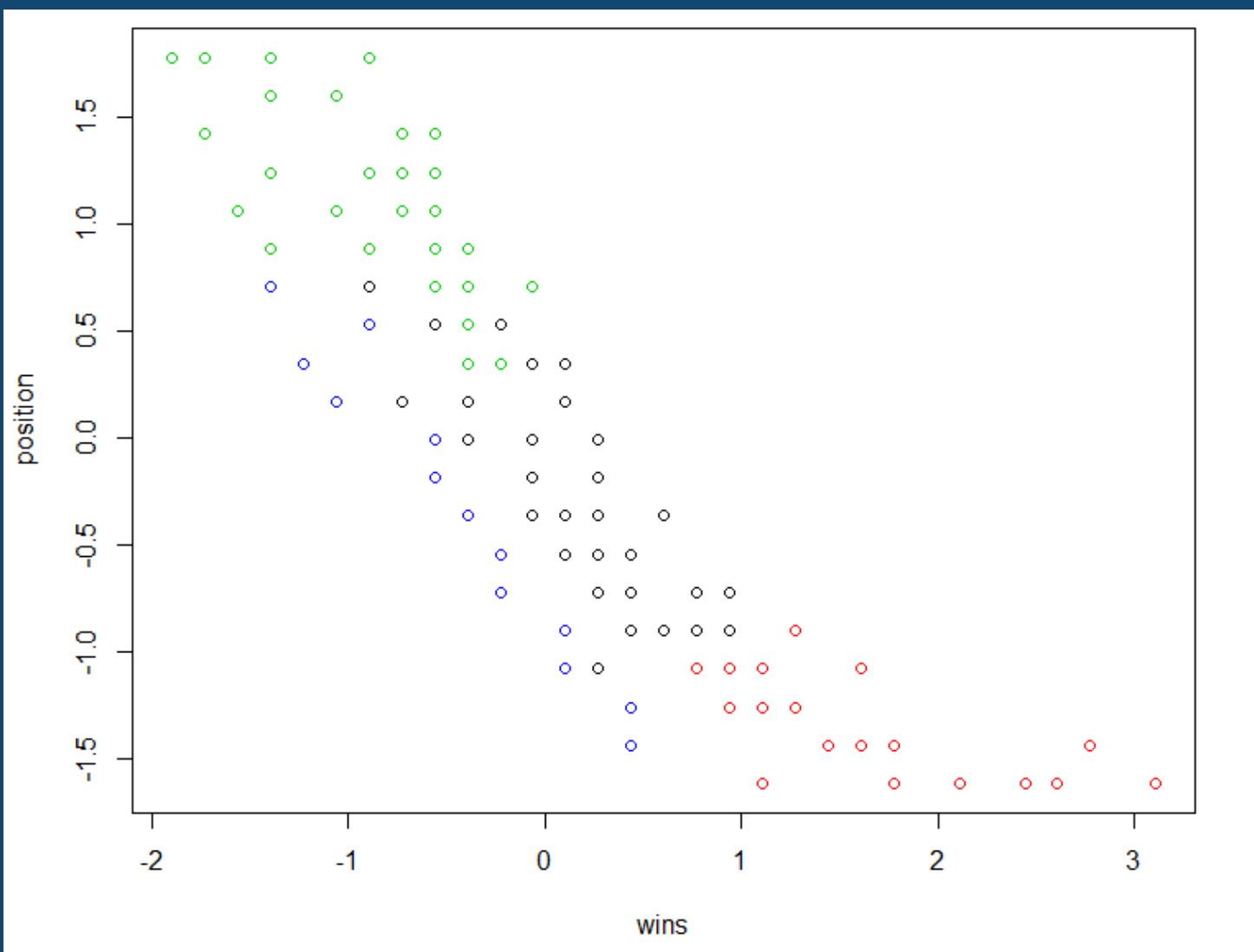
- THERE ARE SOME NEGATIVE COEFFICIENTS IN THE EVALUATION OF THE SILHOUETTE INDEXES
- THIS NUMBER IS LESS THEN THE ONE THAT WE FOUND OUT FOR HIERARCHICAL CLUSTERING
- AVERAGE SILHOUETTE COEFFICIENT IS EQUAL TO 0.3397639
- GOOD RESULTS!



CUSPLOT 2D REPRESENTATION WITH K=4

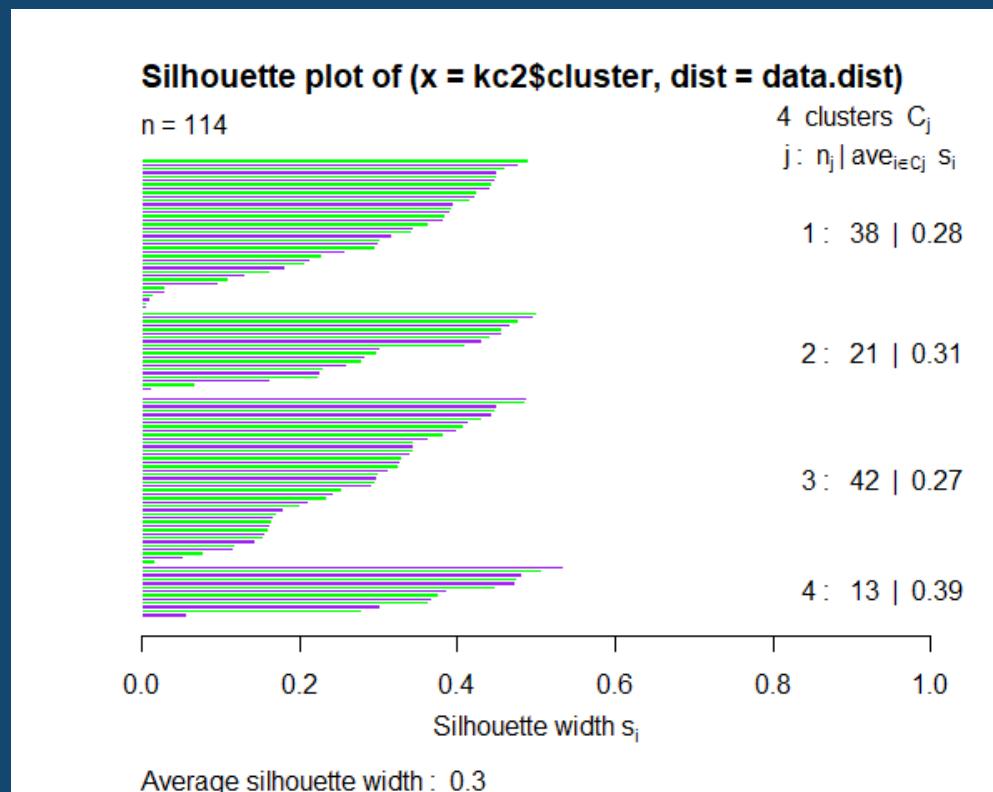


EXAMPLE OF DATA VISUALIZATION WITH K=4



SILHOUETTE COEFFICIENTS FOR K=4

- THERE ARE NO NEGATIVE COEFFICIENTS IN THE EVALUATION OF THE SILHOUETTE INDEXES
- AVERAGE SILHOUETTE COEFFICIENT IS EQUAL TO 0.3
- ALSO HERE GOOD RESULTS!



FINAL CONCLUSIONS ON WHICH TEAM WOULD YOU BET ON?

- THERE ARE 4 CLUSTERS
- THE FOURTH GROUP IS THE ONE WITH BEST WINNING STATISTIC BETS
- THE FIRST IS THE ONE WITH WORST STATISTIC BETS

FURTHER DEVELOPMENTS:

TRY ANALYSIS ON THE WHOLE ORIGINAL DATASET FOR MORE ACCURATE AND TECHNICAL TIPS

Chievo	0	0	1	0
Cordoba	0	0	0	0
Crotone	0	0	0	0
Crystal Palace	1	0	0	0
CSKA Moscow	0	0	0	1
Darmstadt	0	0	0	0
Deportivo La Coruna	0	0	0	0
Dijon	0	0	1	0
Dinamo Moscow	0	0	0	1
Eibar	1	0	0	0
Eintracht Frankfurt	1	0	0	0
Elche	0	0	0	0
Empoli	0	0	1	0
Espanyol	1	0	0	0
Everton	1	0	0	0
Evian Thonon Gaillard	0	0	0	0
FC Cologne	0	0	0	0
FC Krasnodar	0	0	0	1
FC Orenburg	0	0	0	1
FC Rostov	0	0	0	1
FC Ufa	0	0	0	1
FC Yenisey Krasnoyarsk	0	0	1	0
Fiorentina	1	0	0	0
FK Akhmat	0	0	0	1
Fortuna Duesseldorf	0	0	1	0
Freiburg	0	0	1	0

**THANK YOU FOR THE
ATTENTION...**

...AND GOOD LUCK!