



Football - Diversity and Cohesion

- **Question:** How is the mix of diversity/cohesion affecting team identity?

- **4 Contestants:**

- **Regular:** Arsenal FC, Tottenham Hot Spur FC
- **Irregular:** Chelsea FC, Newcastle FC

- **6 Attributes:**

- **Diversity:** Age_Diversity, Club_Diversity, NationalTeam_Diversity
- **Cohesion:** Current_Cohesion, Past_Cohesion, NationalTeam_Cohesion

Numerical Data

Arsenal

- Position

arsenal: [3, 4, 3, 4, 4, 3, 2, 4, 6, 5]

- Diversity

'age_div': [9, 8, 9, 9, 9, 10, 12, 10, 11, 10]

'club_div': [15, 17, 25, 26, 25, 32, 30, 21, 24, 26]

'national_div': [10, 10, 12, 11, 8, 9, 10, 7, 10, 12]

- Cohesion

'curr_16': [1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]

'curr_14': [1.0, 2.0, 1.0, 1.0, 2.0, 1.0, 2.0, 1.0, 1.0, 1.0]

'curr_12': [2.0, 3.0, 1.0, 2.0, 2.0, 1.0, 2.0, 2.0, 2.0, 1.0]

'curr_10': [2.0, 3.0, 2.0, 2.0, 3.0, 1.0, 2.0, 3.0, 2.0, 1.0]

'curr_8': [3.0, 4.0, 3.0, 3.0, 4.0, 2.0, 3.0, 3.0, 4.0, 2.0]

'past_co': [0, 0, 0, 1, 2, 2, 2, 0, 0, 6]

'national_co': [3, 7, 4, 6, 23, 16, 17, 13, 25, 6]

Tottenham

- Position

tottenham: [4, 5, 4, 5, 6, 5, 3, 2, 3, 4]

- Diversity

'age_div': [11, 9, 9, 10, 10, 10, 8, 10, 9, 10]

'club_div': [23, 22, 23, 29, 35, 25, 29, 19, 21, 19],

'national_div': [7, 7, 4, 6, 9, 9, 8, 7, 8, 8],

- **Cohesion**

```
'curr_16': [1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0],  
'curr_14': [1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 3.0],  
'curr_12': [1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 1.0, 2.0, 3.0, 3.0],  
'curr_10': [2.0, 2.0, 2.0, 2.0, 2.0, 2.0, 2.0, 2.0, 3.0, 4.0],  
'curr_8': [2.0, 3.0, 2.0, 2.0, 2.0, 2.0, 3.0, 3.0, 4.0, 4.0],  
  
'past_co': [3, 4, 3, 0, 3, 3, 5, 6, 6, 4],  
'national_co': [9, 13, 13, 14, 16, 18, 15, 15, 16, 26]
```

Chelsea

- **Position**

```
chelsea: [1, 2, 6, 3, 3, 1, 10, 1, 5, 3]
```

- **Diversity**

```
'age_div': [10, 12, 9, 10, 9, 10, 9, 9, 12, 10],  
'club_div': [11, 15, 18, 25, 33, 31, 29, 32, 28, 30],  
'national_div': [11, 11, 9, 9, 10, 9, 8, 7, 8, 9],
```

- **Cohesion**

```
'curr_16': [1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0],  
'curr_14': [1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0],  
'curr_12': [1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 2.0, 1.0, 1.0],  
'curr_10': [1.0, 1.0, 1.0, 2.0, 2.0, 1.0, 2.0, 3.0, 2.0, 2.0],  
'curr_8': [2.0, 2.0, 1.0, 2.0, 2.0, 2.0, 2.0, 3.0, 2.0, 3.0],  
  
'past_co': [0, 0, 0, 0, 0, 4, 6, 8, 5, 4],
```

'national_co': [3, 4, 7, 12, 17, 19, 25, 22, 18, 19]

Newcastle

- Position

newcastle: [21, 12, 5, 16, 10, 15, 18, 21, 10, 13]

- Diversity

'age_div': [10, 11, 10, 8, 10, 9, 9, 9, 9, 6],

'club_div': [14, 23, 25, 24, 20, 25, 21, 26, 43, 35],

'national_div': [4, 7, 7, 6, 6, 7, 9, 7, 10, 12],

- Cohesion

'curr_16': [1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0],

'curr_14': [1.0, 1.0, 1.0, 2.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0],

'curr_12': [1.0, 1.0, 1.0, 2.0, 1.0, 1.0, 1.0, 1.0, 2.0, 1.0],

'curr_10': [2.0, 2.0, 1.0, 2.0, 2.0, 2.0, 2.0, 1.0, 2.0, 1.0],

'curr_8': [2.0, 2.0, 2.0, 3.0, 2.0, 2.0, 2.0, 2.0, 2.0, 3.0],

'past_co': [0, 3, 0, 0, 9, 2, 2, 3, 1, 5],

'national_co': [6, 7, 9, 16, 9, 7, 11, 7, 6, 3]

The Reason for Curr_16 to Curr_8

We are talking about how many seasons that the current players have played for the team. Therefore, there could be a problem: if there is a new player in each season (which is very possible), then current_CO will always be only 1 season.

Therefore, additionally, I set curr_14 to curr_8 so that we can remove possible influences of some new players and see the current_CO among old players in the team. For example, curr_8 means that we remove 8 newest players and see how many seasons that remaining players have played together for the team.

The Reason for Z-Score Normalization

Different data could be in different scales. For example, position and age_diversity are in different dimensions. Therefore, when you want to compare different data, data normalization should be applied to remove different scales, so that data could be compared together. And z-score is one particular normalization approach.

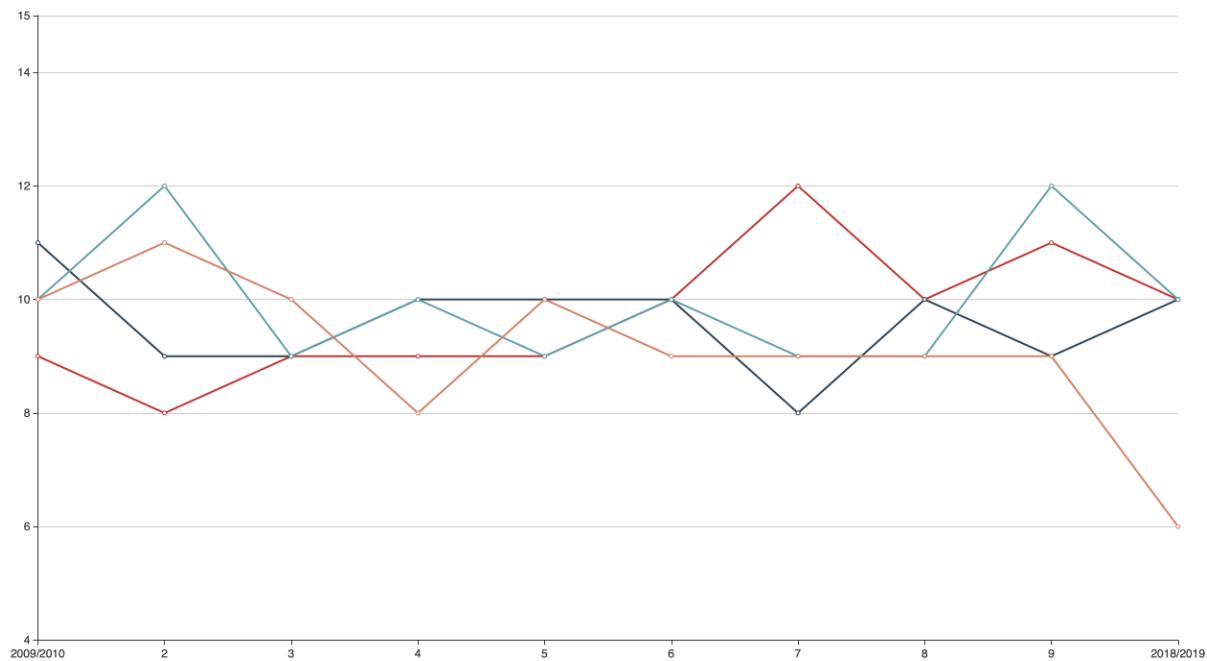
The data processed with z-score is normally distributed and non-scaled, thus is better for comparison, especially in charts.

Diversity Comparison Among 4 Teams

Age Diversity (number)

Age Diversity Comparison

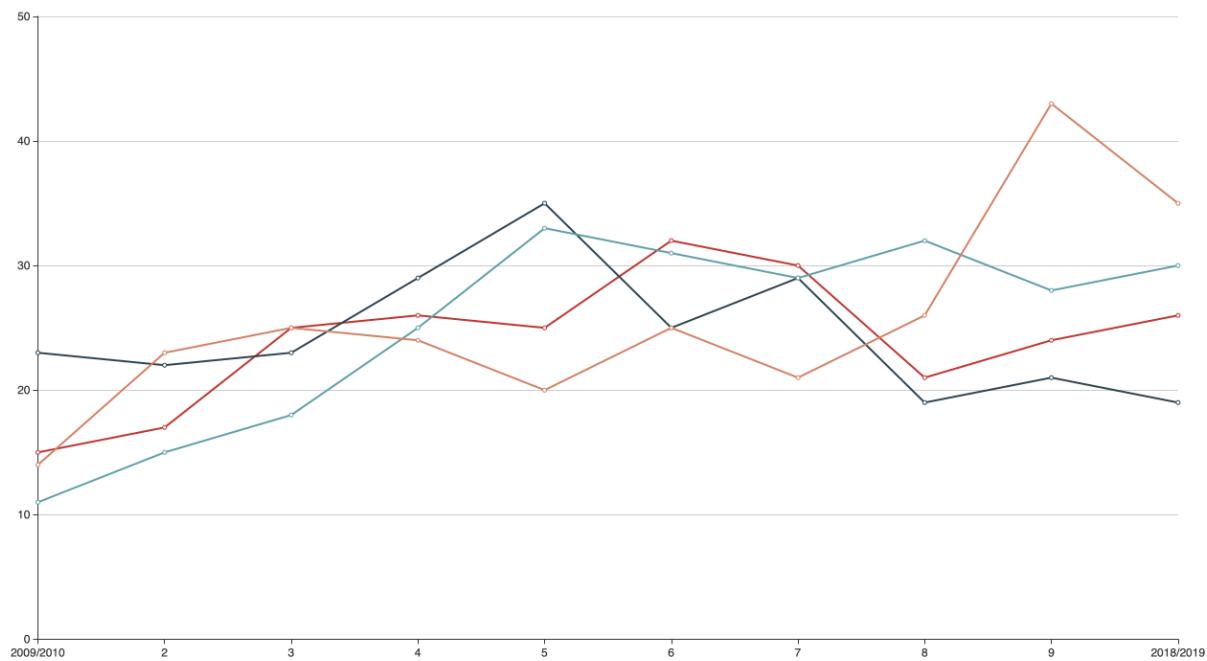
-○- Arsenal -○- Tottenham -○- Chelsea -○- Newcastle



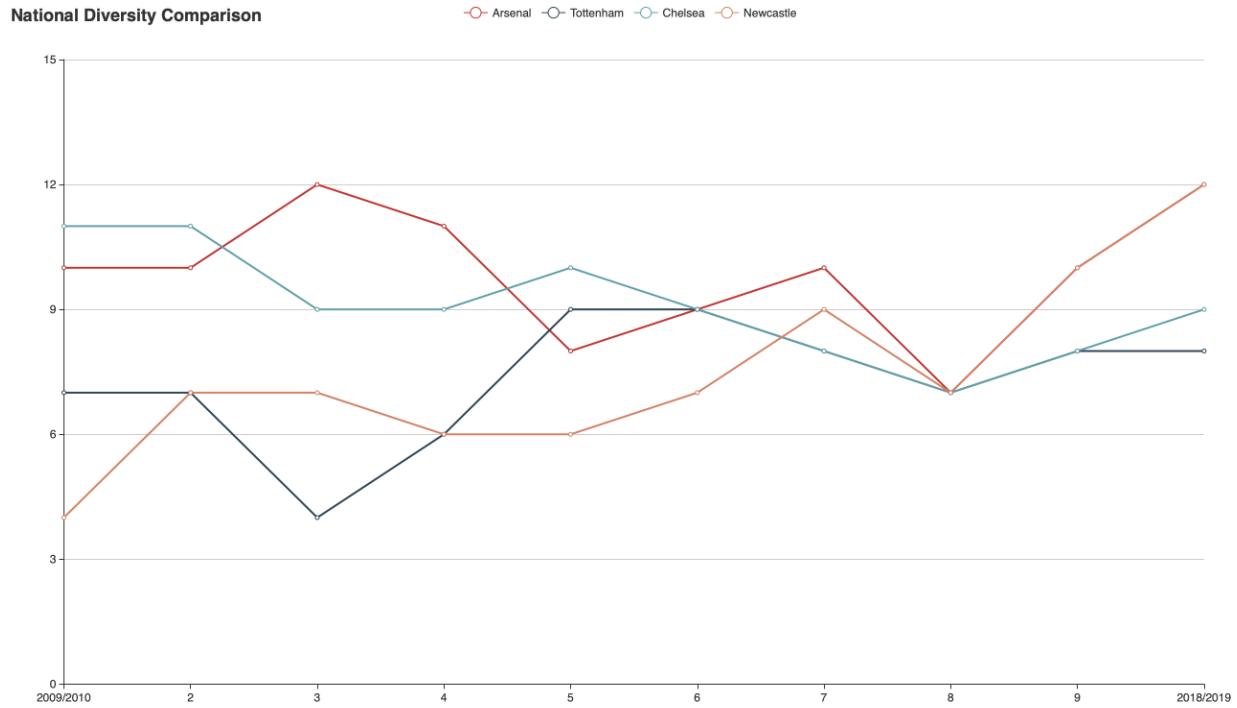
Club Diversity (number)

Club Diversity Comparison

-○- Arsenal -○- Tottenham -○- Chelsea -○- Newcastle



National Team Diversity (number)

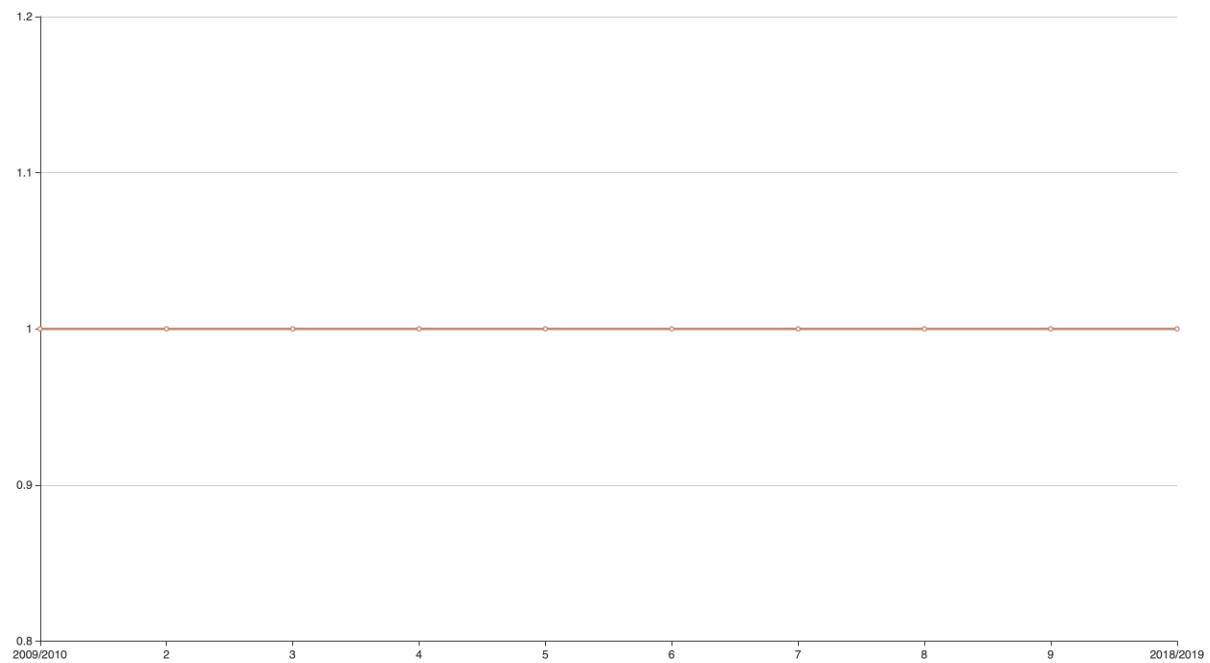


Cohesion Comparison Between 4 Teams

Current Cohesion (all 16 players)

Current Cohesion (16) Comparison

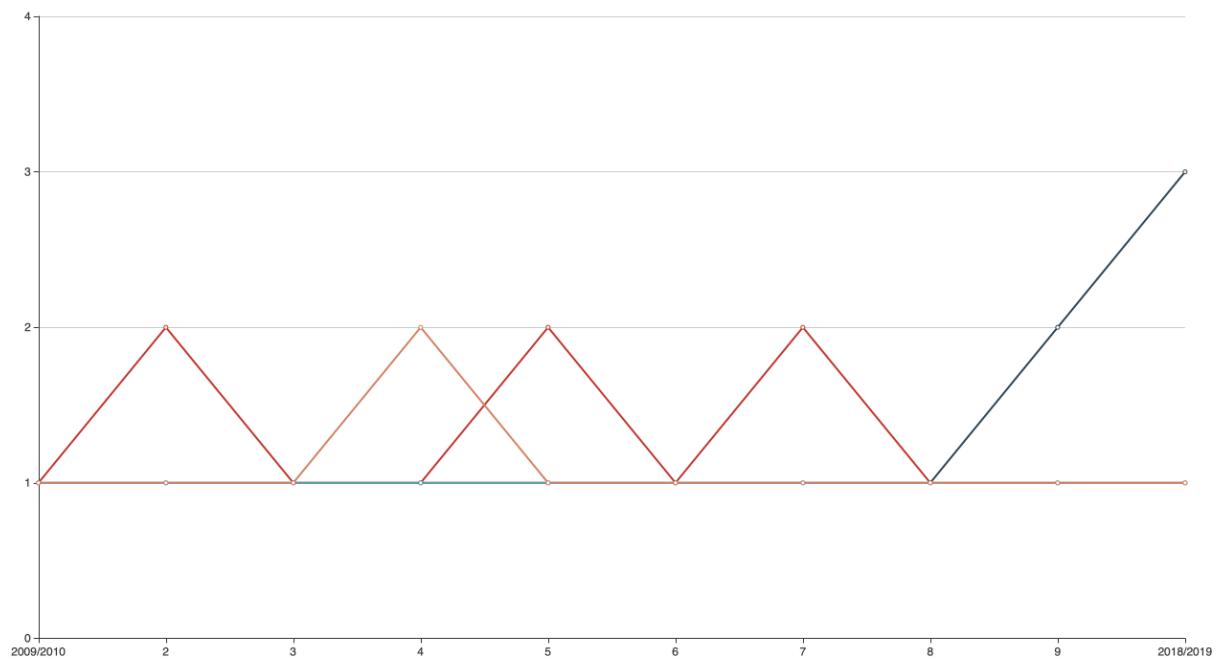
—○— Arsenal —○— Tottenham —○— Chelsea —○— Newcastle



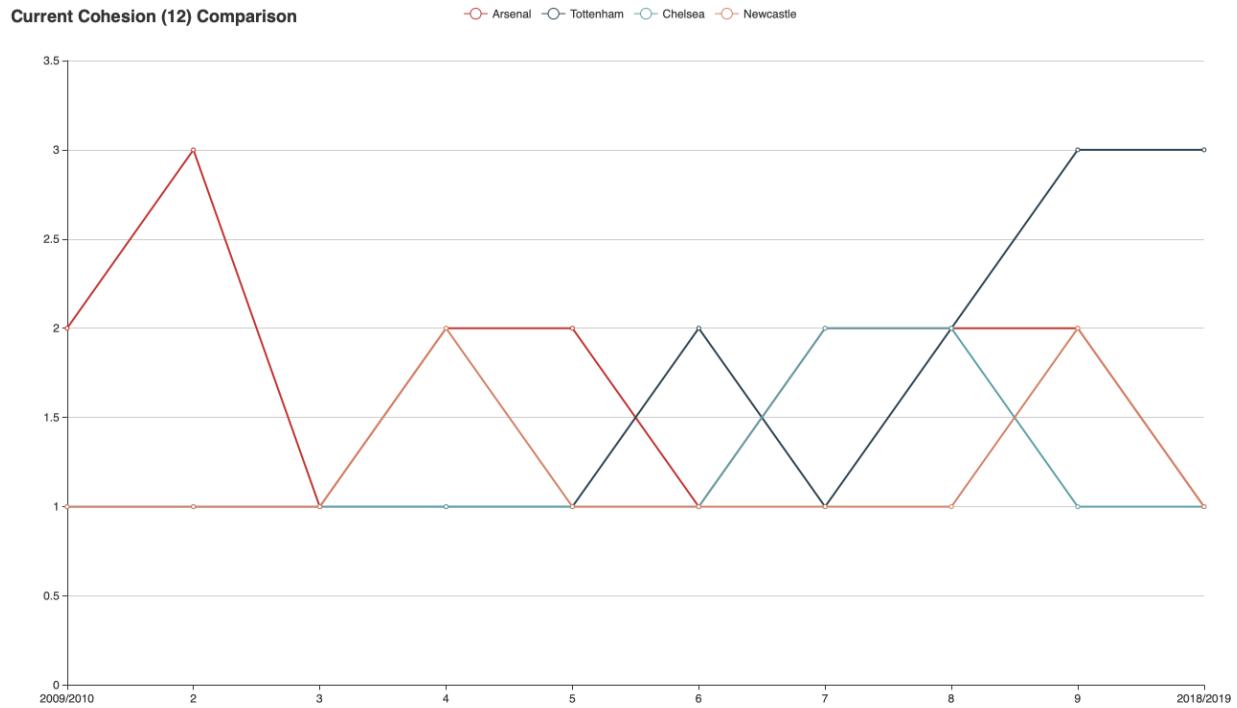
Current Cohesion (14 longest players)

Current Cohesion (14) Comparison

—○— Arsenal —○— Tottenham —○— Chelsea —○— Newcastle



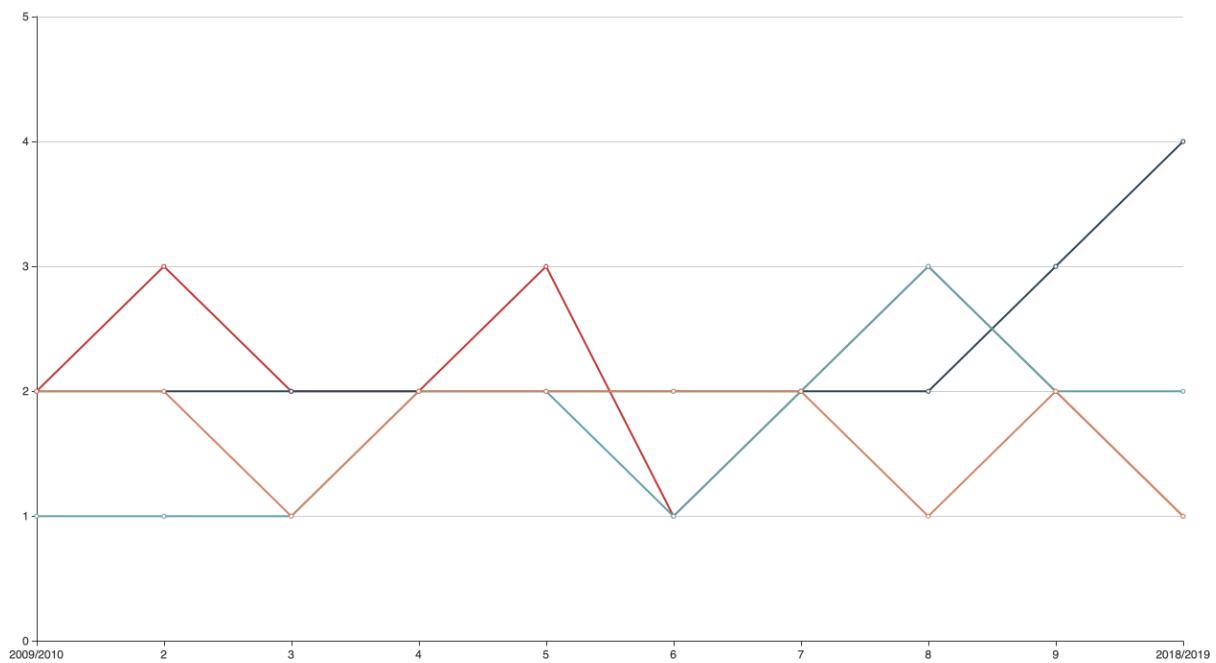
Current Cohesion (12 longest players)



Current Cohesion (10 longest players)

Current Cohesion (10) Comparison

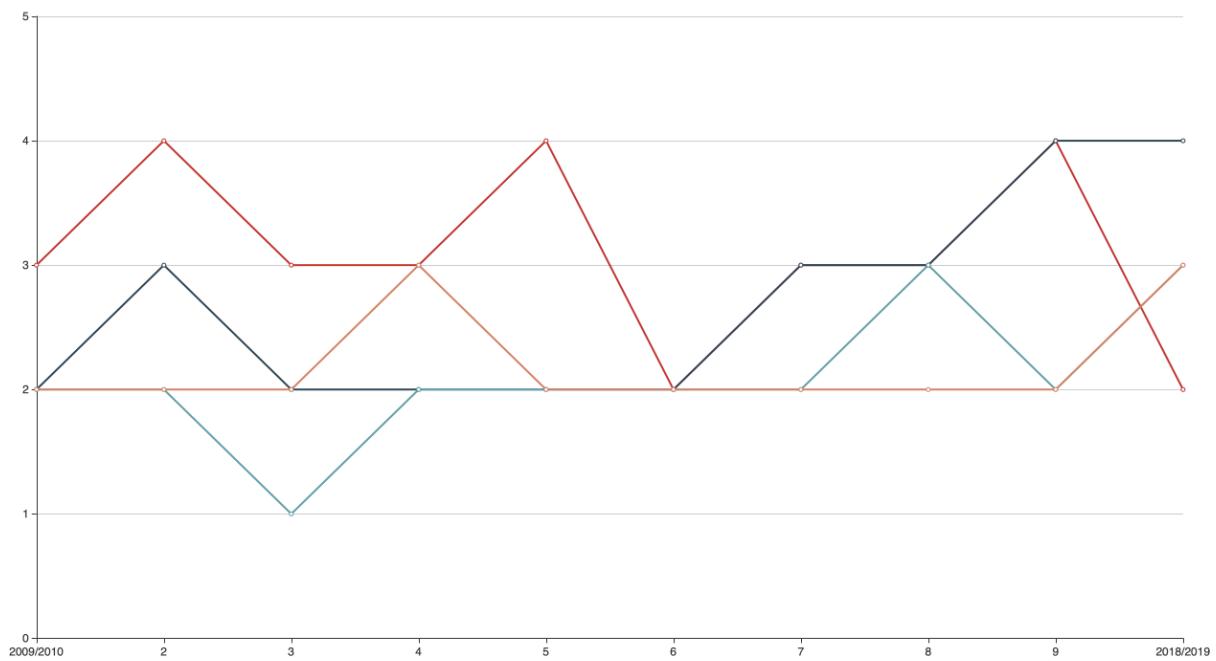
-○- Arsenal -○- Tottenham -○- Chelsea -○- Newcastle



Current Cohesion (8 longest players)

Current Cohesion (8) Comparison

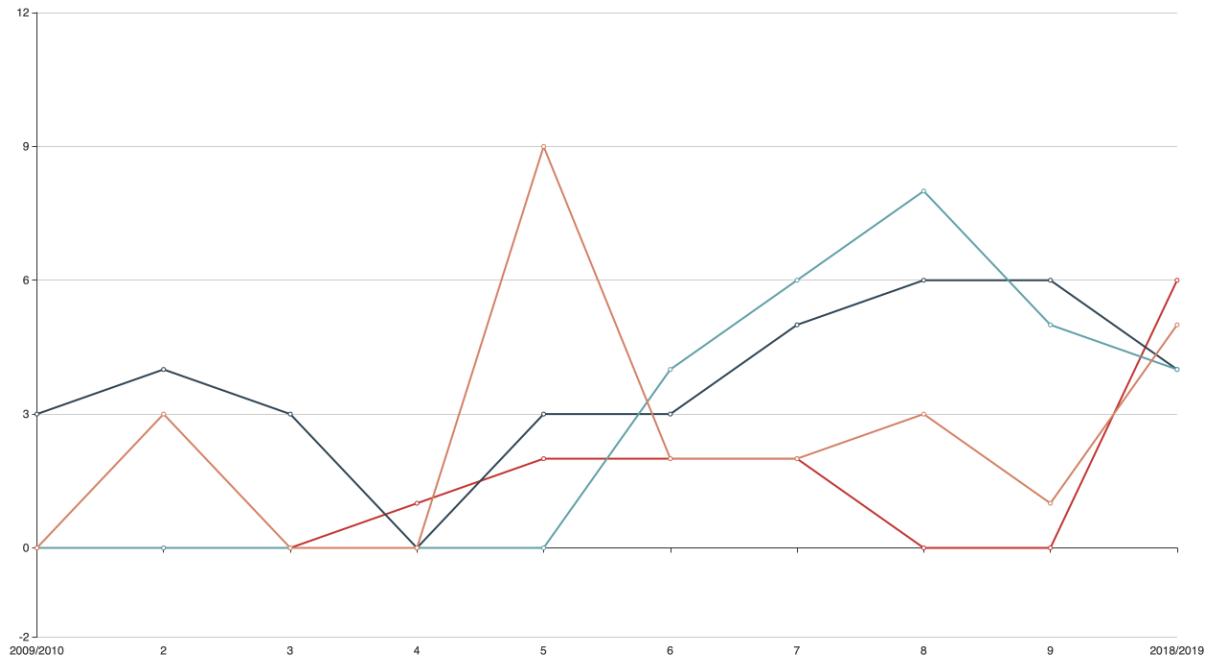
-○- Arsenal -○- Tottenham -○- Chelsea -○- Newcastle



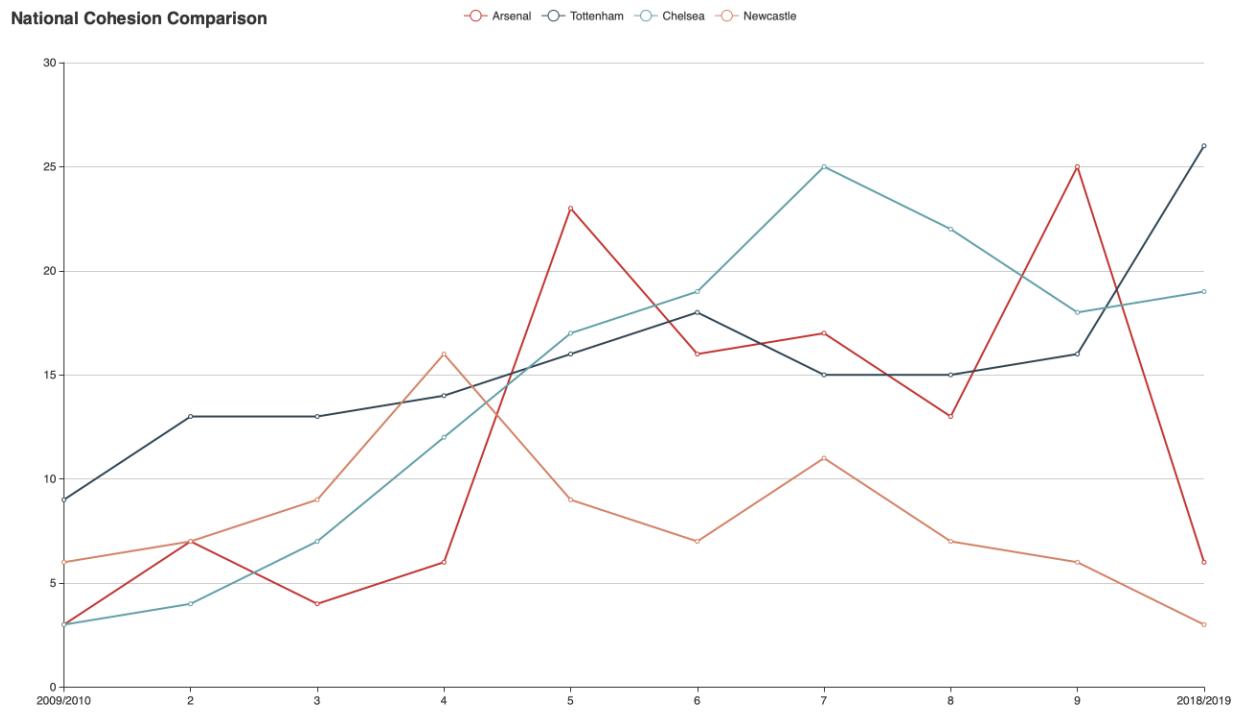
Past Cohesion (Season)

Past Cohesion Comparison

-○- Arsenal -○- Tottenham -○- Chelsea -○- Newcastle



National Cohesion (Season)



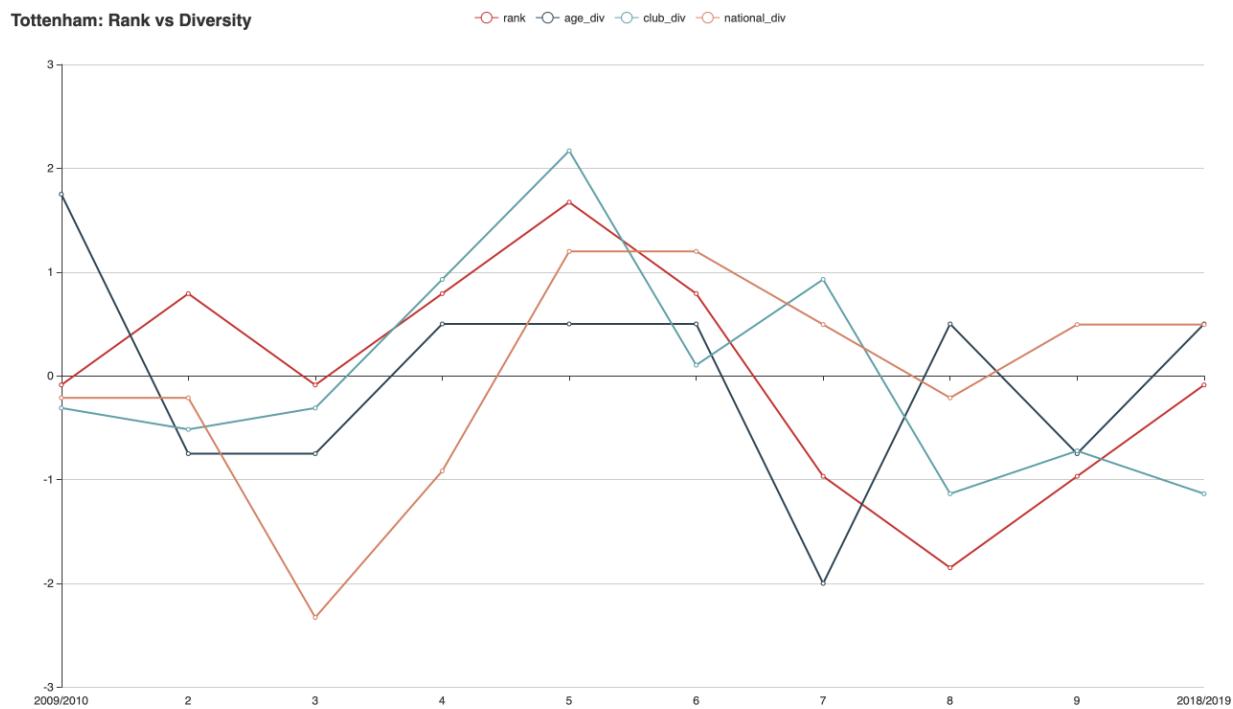
Position vs Diversity for Four Teams

| With Z-Score Normalization

Arsenal

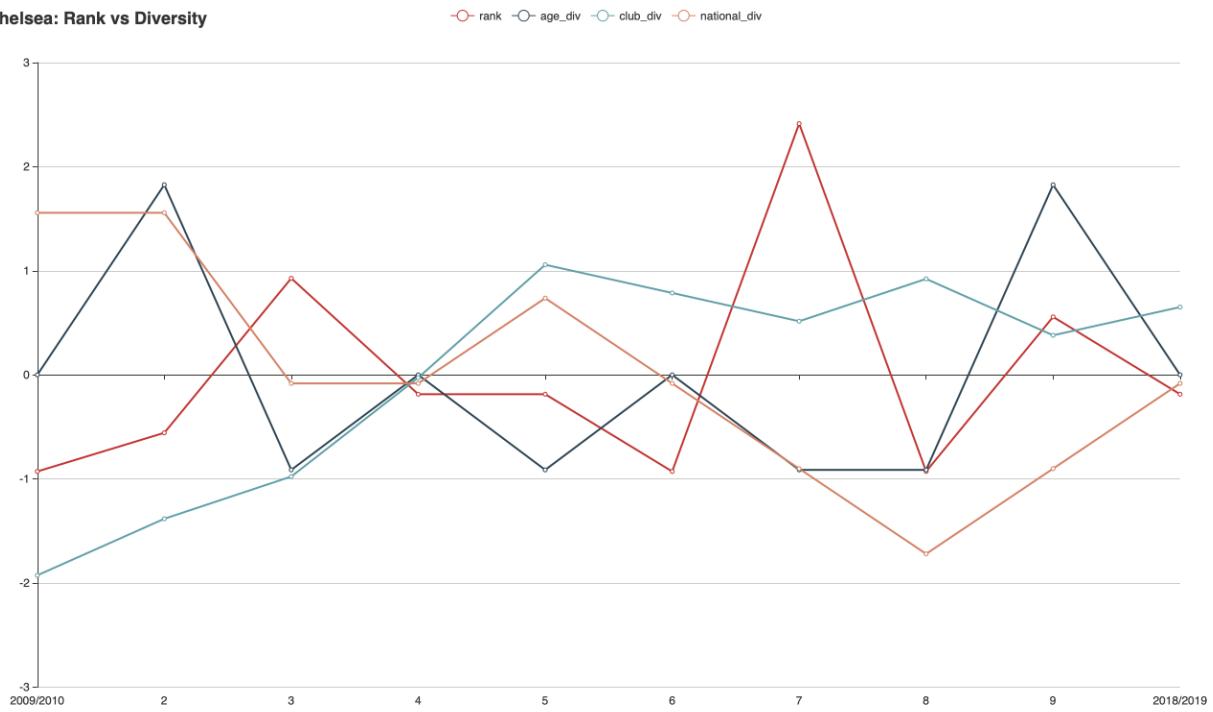


Tottenham

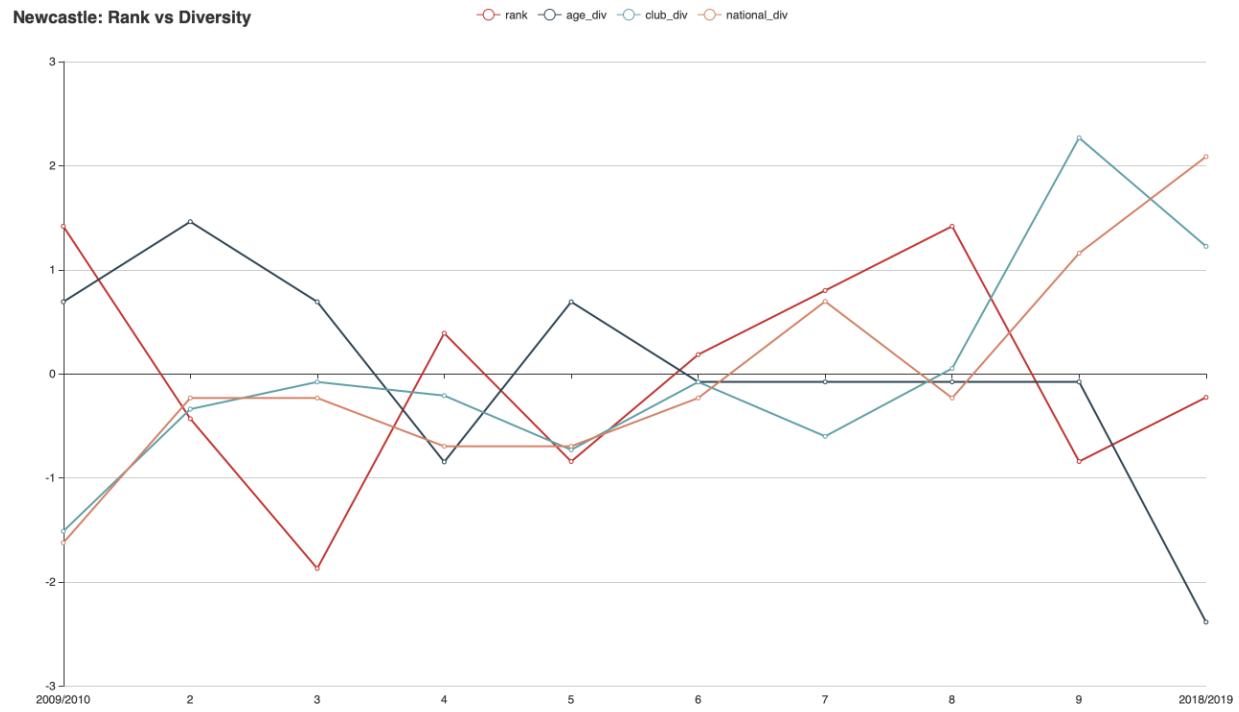


Chelsea

Chelsea: Rank vs Diversity



Newcastle

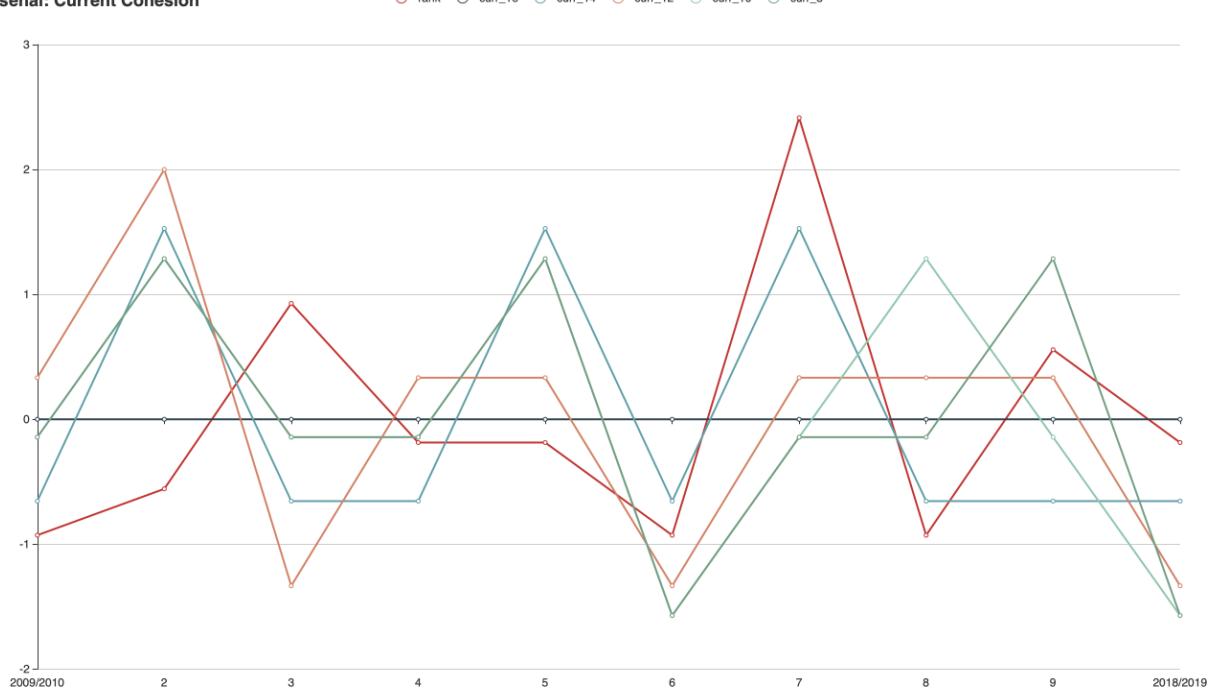


Position vs Current Cohesions for Four Teams

| With Z-Score Normalization

Arsenal

Asenal: Current Cohesion



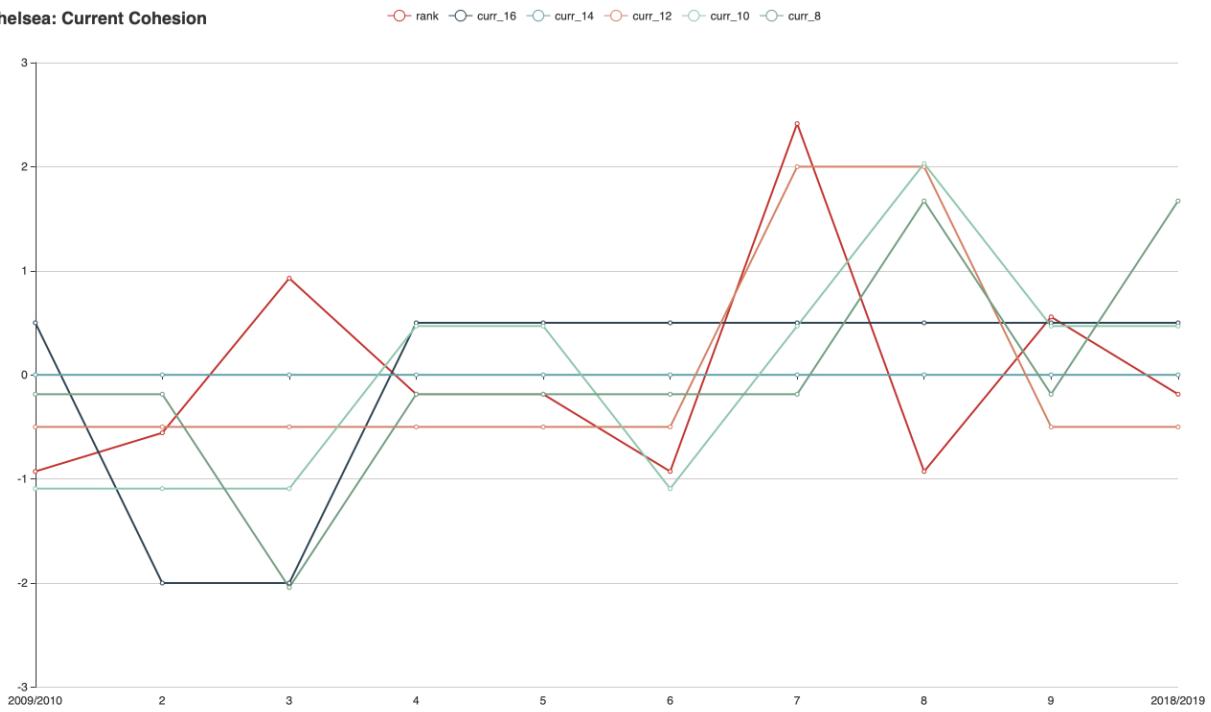
Tottenham

Tottenham: Current Cohesion

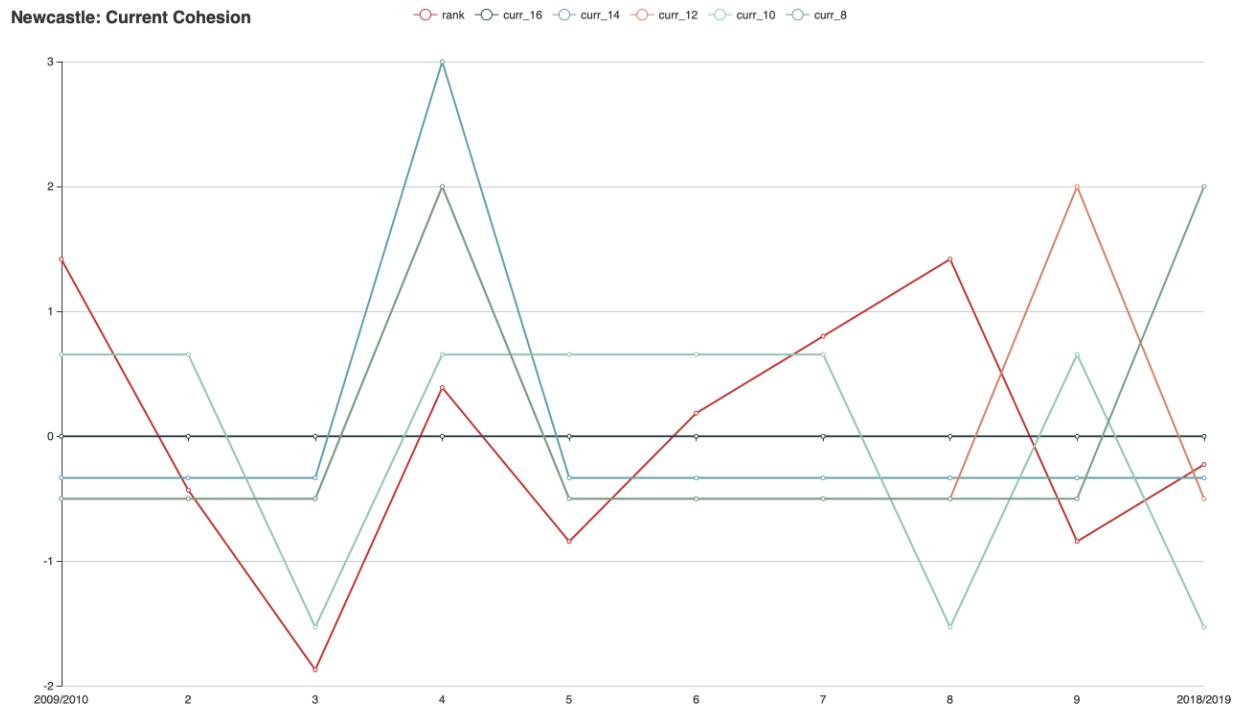


Chelsea

Chelsea: Current Cohesion



Newcastle



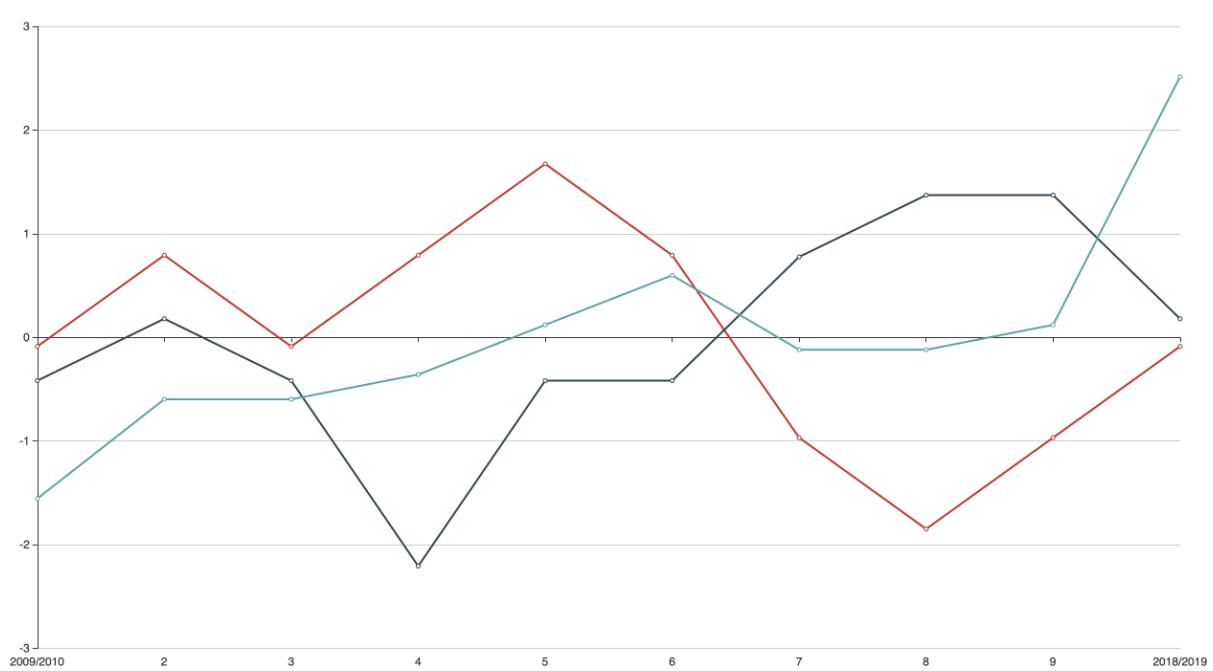
Position vs Club and National Cohesions for Four Teams

| With Z-Score Normalization

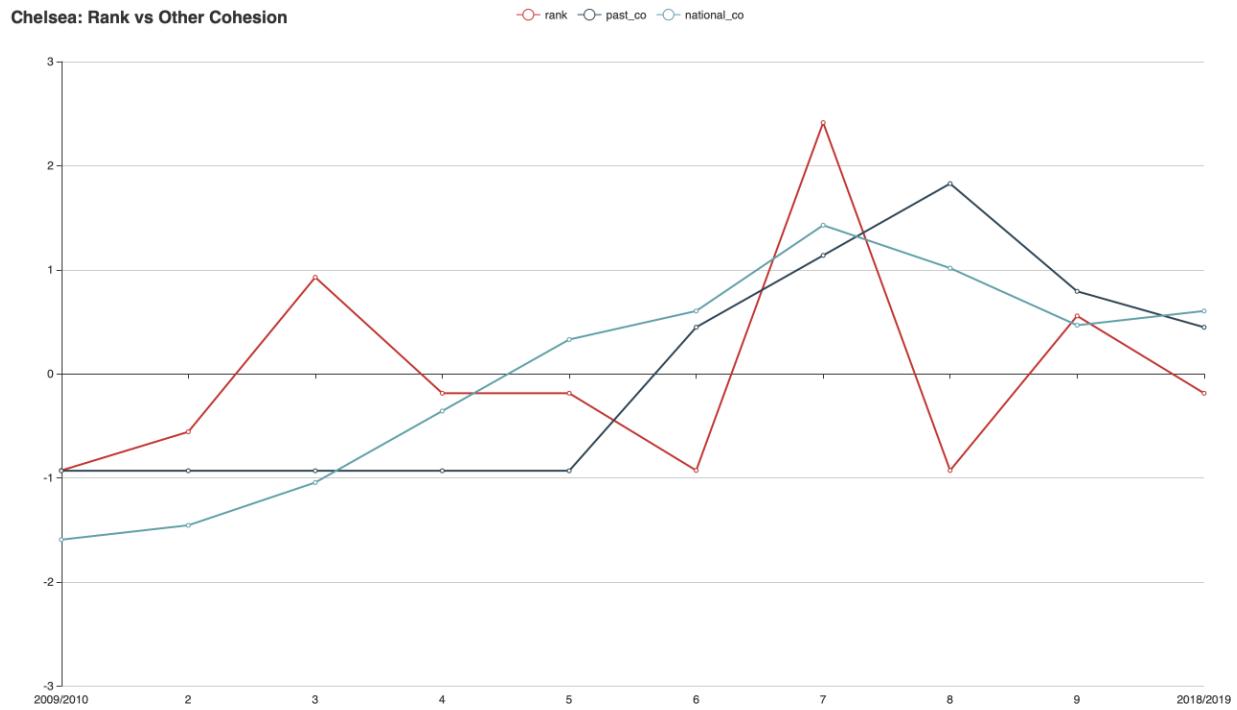
Arsenal

Arsenal: Rank vs Other Cohesion

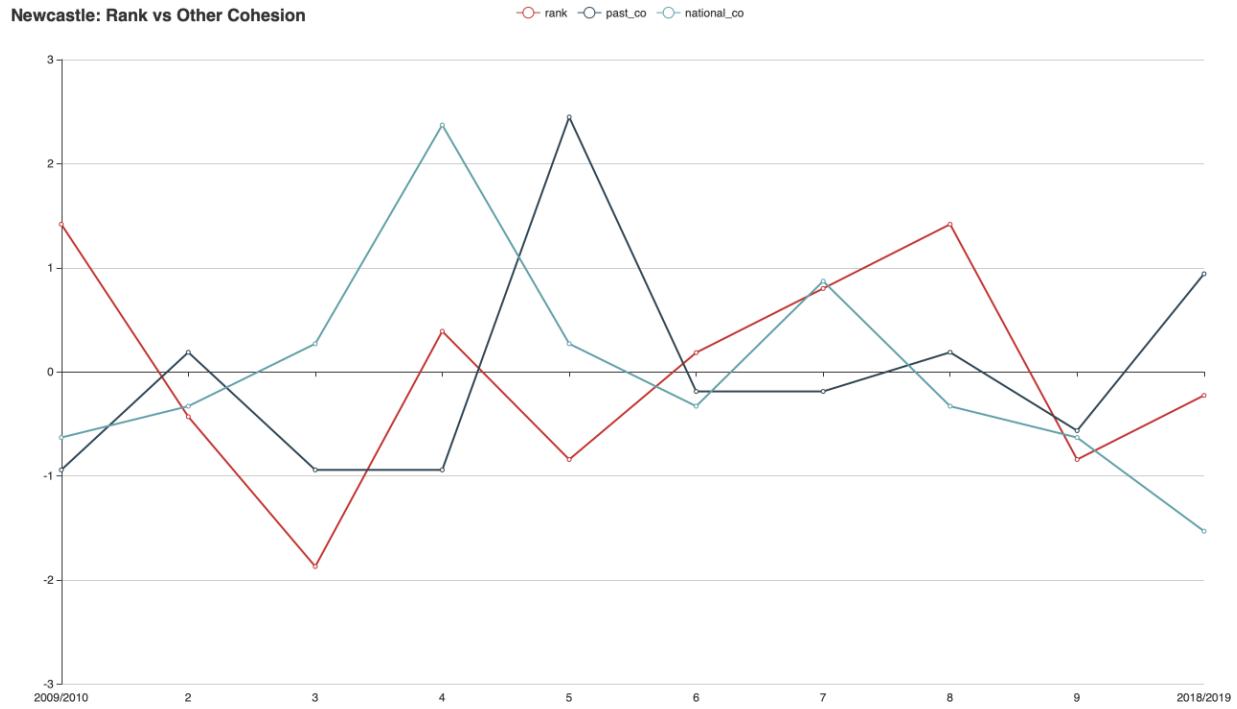
Tottenham

Tottenham: Rank vs Other Cohesion

Chelsea



Newcastle



General Conclusions

1. It's difficult to see team identities through current attributes we chose.
2. Teams with different identities have similar degree of fluctuation with current attributes.
3. Team diversity and cohesion could fluctuate whether the position fluctuates or not.