Premier League 2018-19 Performance Analysis

- Jasmine Zhang (yz4720)
- Yuxin Zhang (yz4719)
- Tianyou He (th3072)
- Xiaoyu Wu(xw2962)
- Xiaoyu Huang (xh2636)

Motivation

As the most popular football league worldwide, the Premier League consists 20 competitive football clubs in England and attracts the most famous football players. Statistical analysis is widely used in order to achieve better game performance for the club. We set our interest on the PL season 2018/19 because it is the season pre-pandemic and due to data availability. We tend to investigate the relationship between transfer market status, player characteristics, match style and the final standings of the season.

Intended Final Products

Our intended final products include different types of plots that show characteristics of the match and players of all 20 Premier League clubs. Evaluation of popular match styles and their correlation with the final standings of the season.

Anticipated Data Resource

https://datahub.io/sports-data/english-premier-league#readme
https://www.kaggle.com/datasets/thesiff/premierleague1819/
https://fbref.com/en/comps/9/2018-2019/stats/2018-2019-Premier-League-Stats#all_stats_standard

```
tranfer_df = read_csv("./data/transfer.csv")
player_df = read_excel("./data/player_game.xlsx", sheet = 1)
match_df = read.csv("./data/match_data.csv")
```

Planned Analyses

- Interactive / shiny map: how winning percentage varies with geographic location of football stadiums
- Line graph: Transfer market spending over the past decade for the PL clubs
- Scatterplot comparison and heatmap: Match style (possession %, short pass, long pass, conversion rate, corner counts)
- Bar plot / boxplot: player distribution by nationality and position on pitch
- Linear model: minutes played vs. age

Planned timeline

- Plot draft before 11/25
- Result analysis before 12/1
- Final product and video before 12/7