

# **DATA SCIENCE PROJECT**

# **ANALYSIS OF**

# **FOOTBALL**

# **MATCHES**

# CONTENT

- 
- 01** MOTIVATION
  - 02** OUR RESEARCH QUESTION
  - 03** ABOUT THE DATA
  - 04** VISUALIZATION, INTERPRETATION & SIGNIFICANCE
  - 05** CONCLUTIONS & DISCUSSION
  - 06** ASSUMPTIONS & LIMITATIONS
  - 07** NEXT UP

# MOTIVATION



Our motivation to work on the football dataset came from our love for the sport and because it has a global fan base.



Football data can be used to answer a variety of questions about the sport. This includes questions about player performance, team strategy, and the overall dynamics of the game.



# OUR RESEARCH QUESTION

**Primary:** Does having more left-footers affect the outcome of football matches in the English Premier League?

**Secondary:** Do other physical properties of (a) player/s affect the outcome of a match?

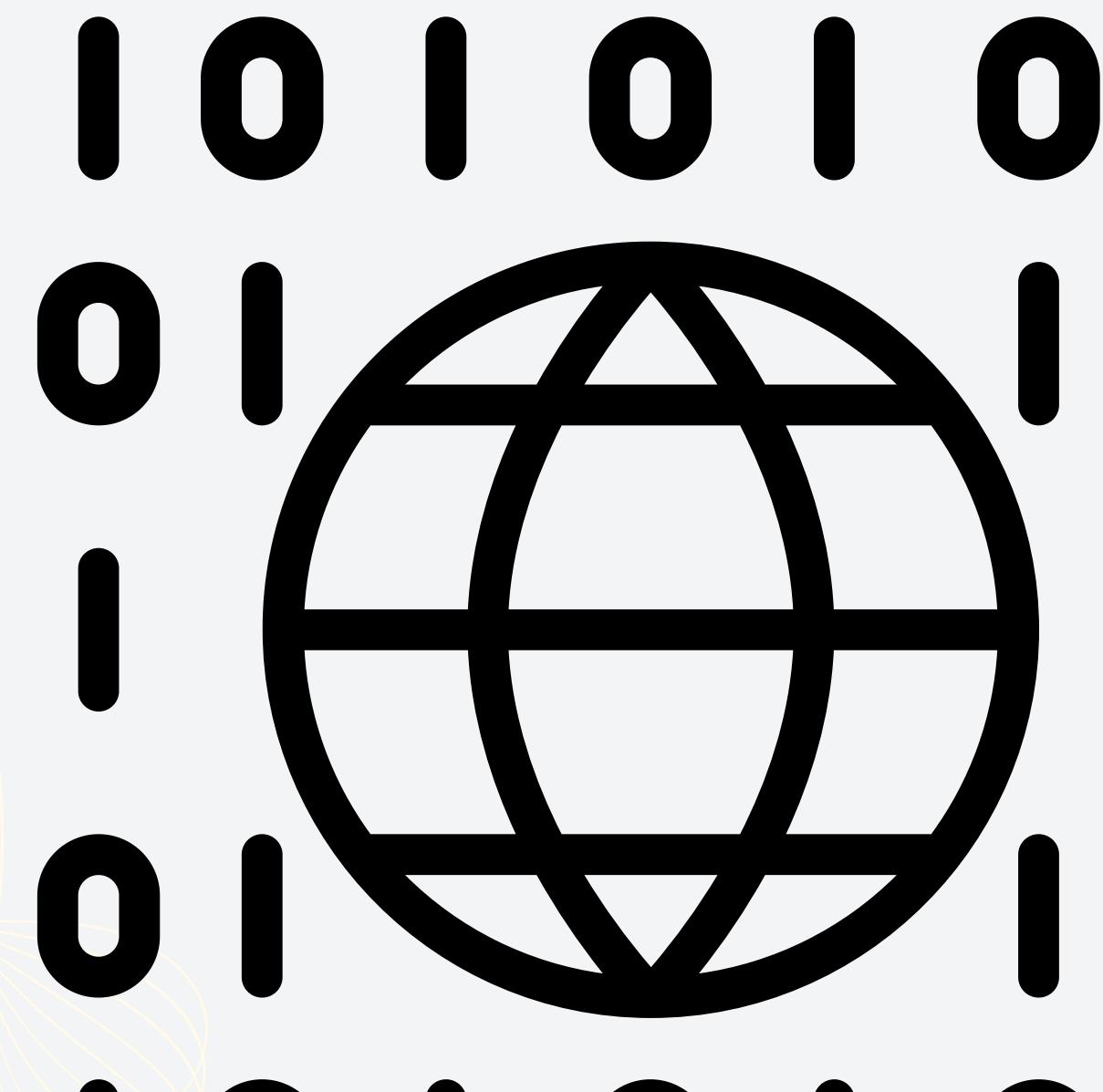
- Height of players (min/max/avg)
- Weight of players (min/max/avg)

**Null Hypothesis -**  
The number of left-footed players does not impact the outcome of the match

# ABOUT THE DATA

The data was from publicly available (spatio-temporal) football data (study published in Nature, Scientific Data) which has metrics from the top five European matches and the World Cup.

Pappalardo, L., Cintia, P., Rossi, A. et al. A public data set of spatio-temporal match events in soccer competitions. Sci Data 6, 236 (2019). <https://doi.org/10.1038/s41597-019-0247-7>



# ABOUT THE DATA

The data was arranged as  
Matches, Events, and  
Individual data.

In Complex nested JSON  
and CSV files

**~40 M**

**Datapoints**



# ABOUT THE DATA

## Matches

each match included season, teams, players for each team, their match highlights, referees details etc.

## Events

Spatiotemporal data including passes, throw-ins, free kicks etc.

## Individual

Details about players, coaches, referees, teams etc.





01

02

03

## EXTRACT

Data had to be **manually extracted** as it was in the form of complex nested JSON and CSV files. Special considerations were needed as **some data was not collected** in **some countries** and **tournaments**

## CLEAN

Somewhat clean. We needed to consider some "0" values and "null" (str) values.

## ANALYSE

Analysis was done in python with the help of specific libraries.

# VISUALIZATION

Proportion of left footers vs right footers

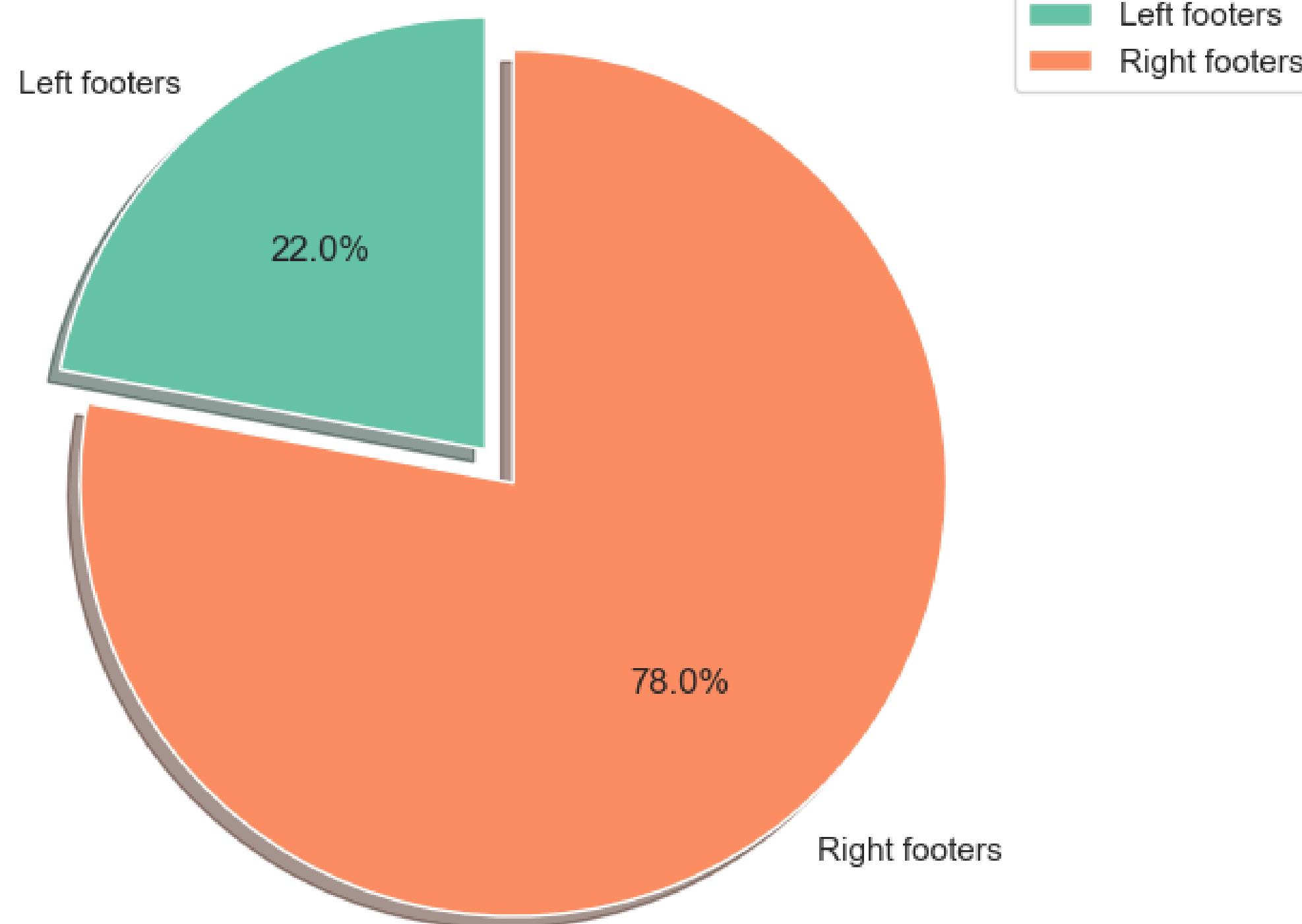
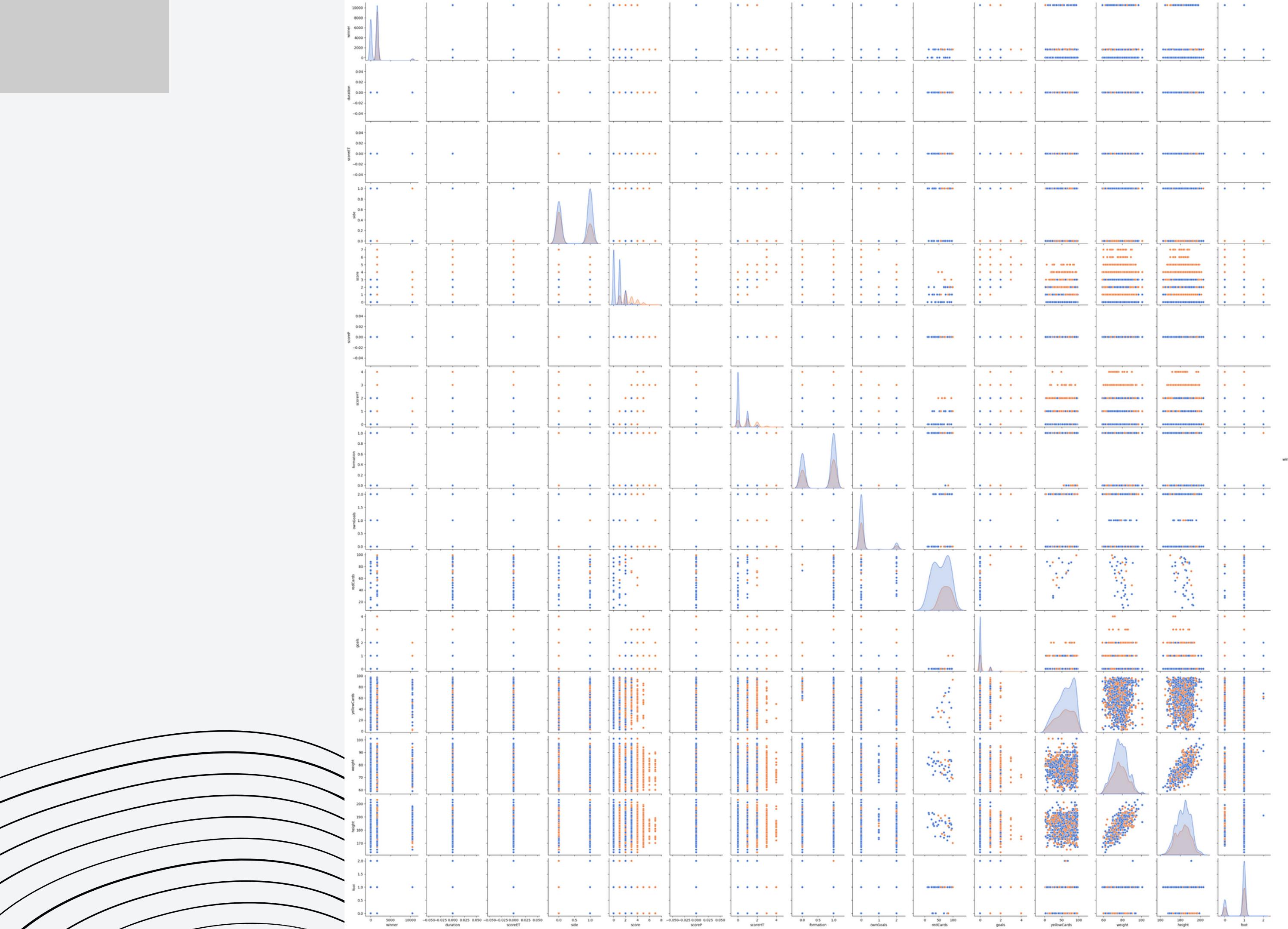
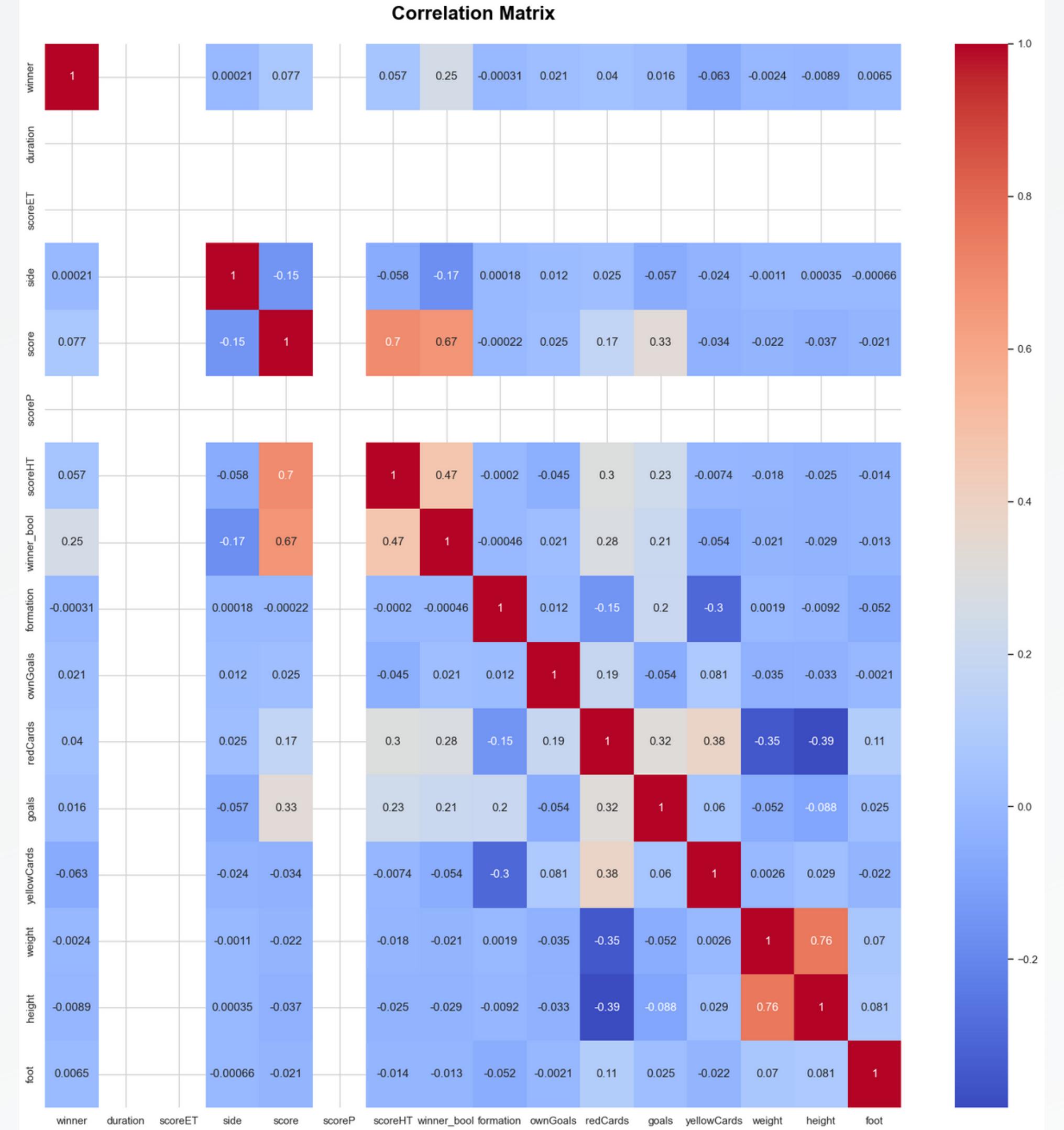


Fig: Footedness of the players in the tournament (England)





## Avg Number of left footers in winning teams vs loosing teams

t\_statistic: 1.44

p value: 0.15

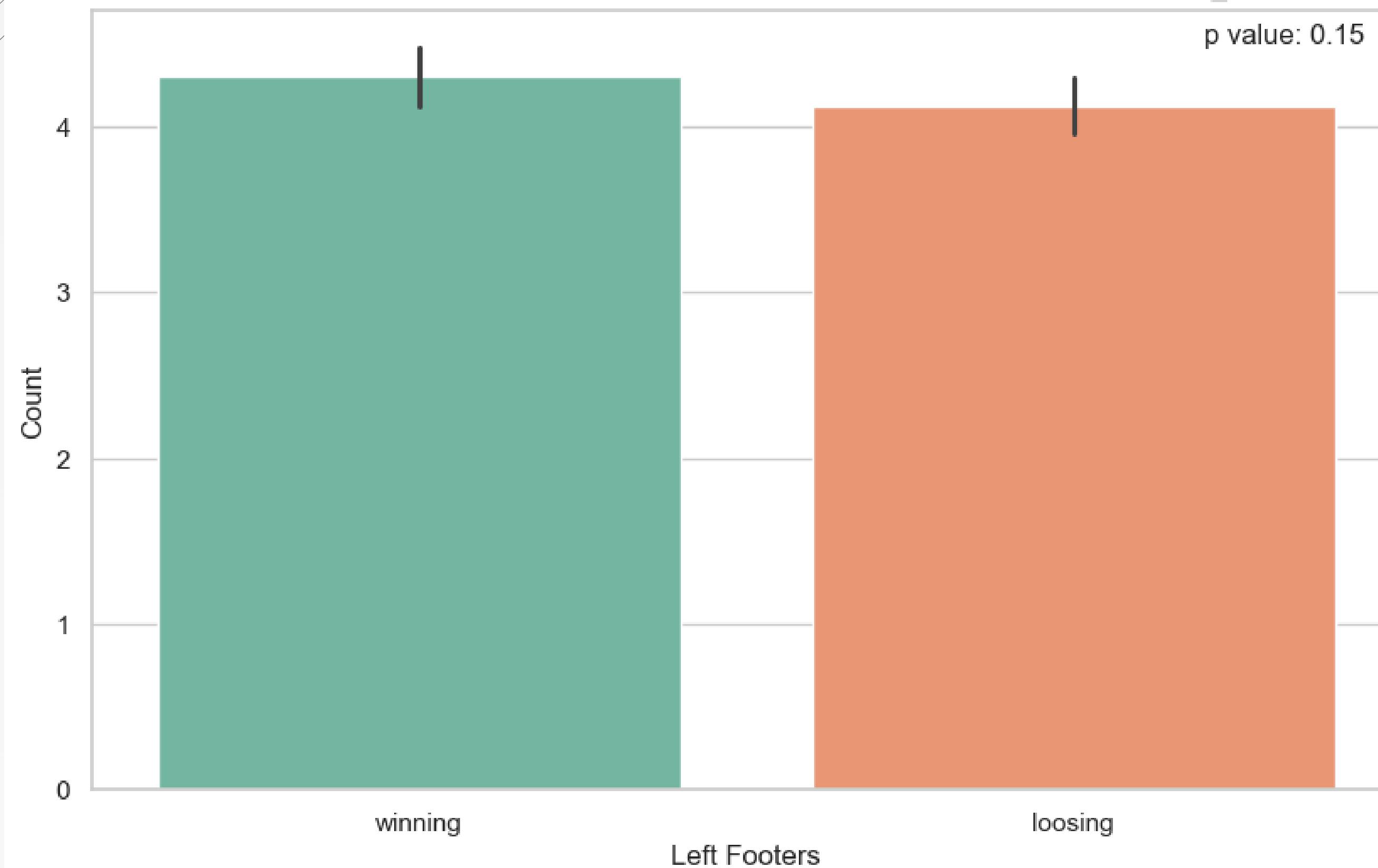


Fig: Average number of left footers in winning vs loosing team

## Difference in number of left footers (winning team - loosing team) vs Number of times won

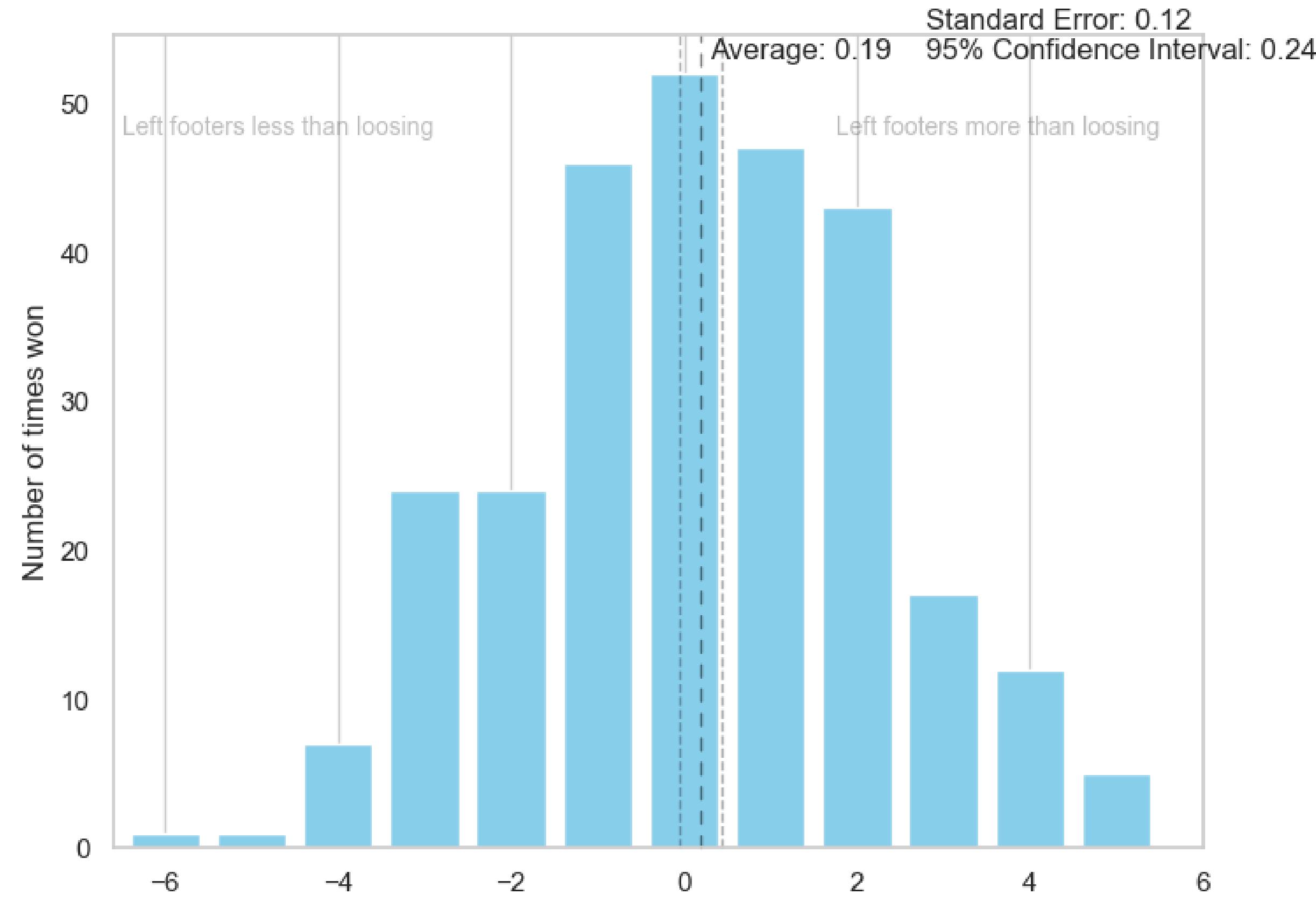


Fig: Distribution between left-footed players and winning

No of Left Footers	Wins	Losses	Draws	Total	Winning Probability
0	0	0	1	1	0.0
1	9	9	6	24	0.375
2	28	30	20	78	0.358974
3	53	56	40	149	0.355705
4	65	77	50	192	0.338542
5	55	58	51	164	0.335366
6	44	31	26	101	0.435644
7	22	16	8	46	0.478261
8	3	2	0	5	0.6

Table: Winning probability with number of left foot players

## Number of left footers vs winning probability

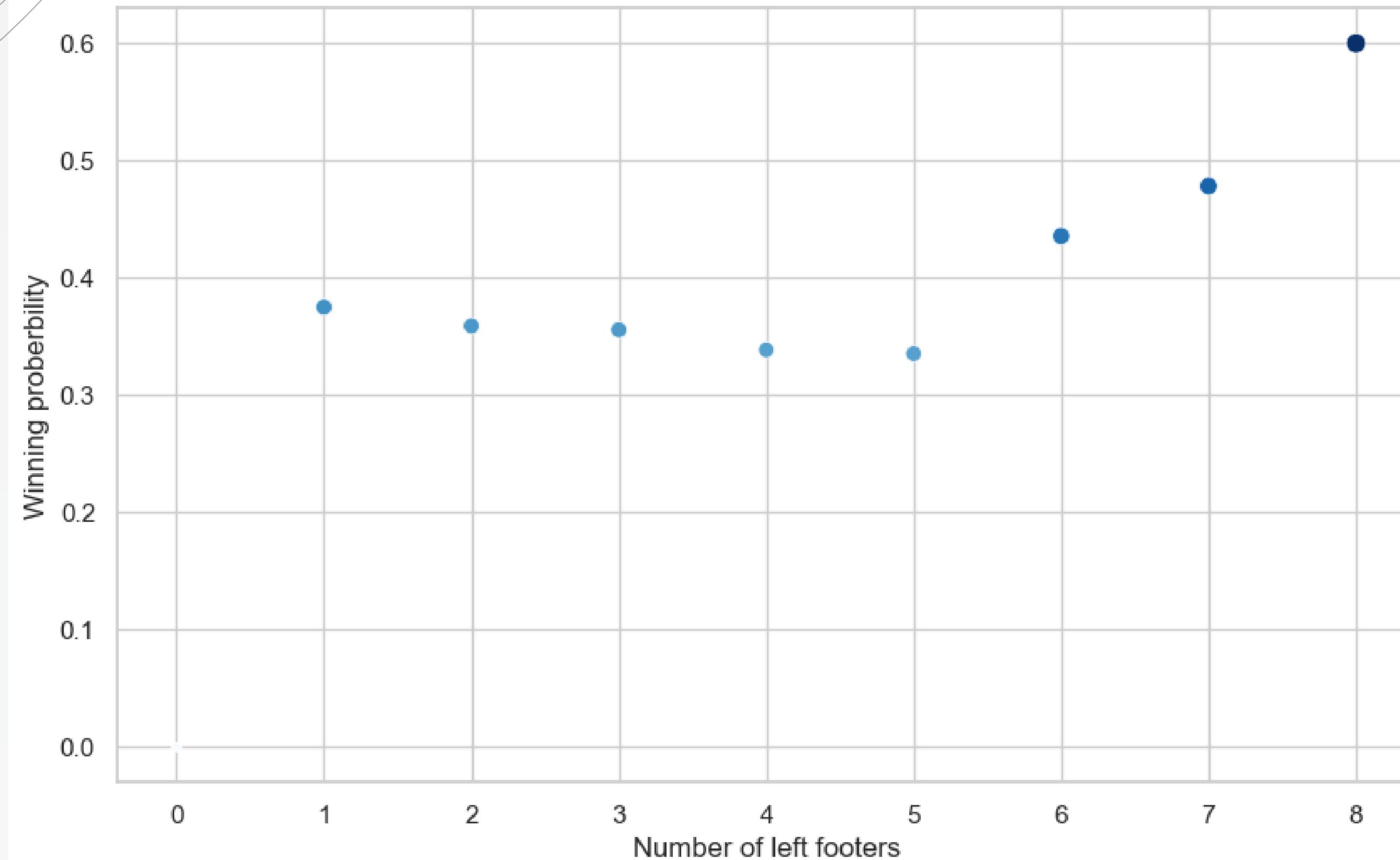


Fig: Scatter plot of winning probability with number of left-footed players

	Winning	Loosing
Left footers	1201	1149
right footers	3807	3847

t- statistic score: 1.28  
p-value: 0.20

Table: Contingency table, left & right footed players played matchup in winning and losing teams

## Left-footer distribution in winning and loosing teams

p value: 0.058

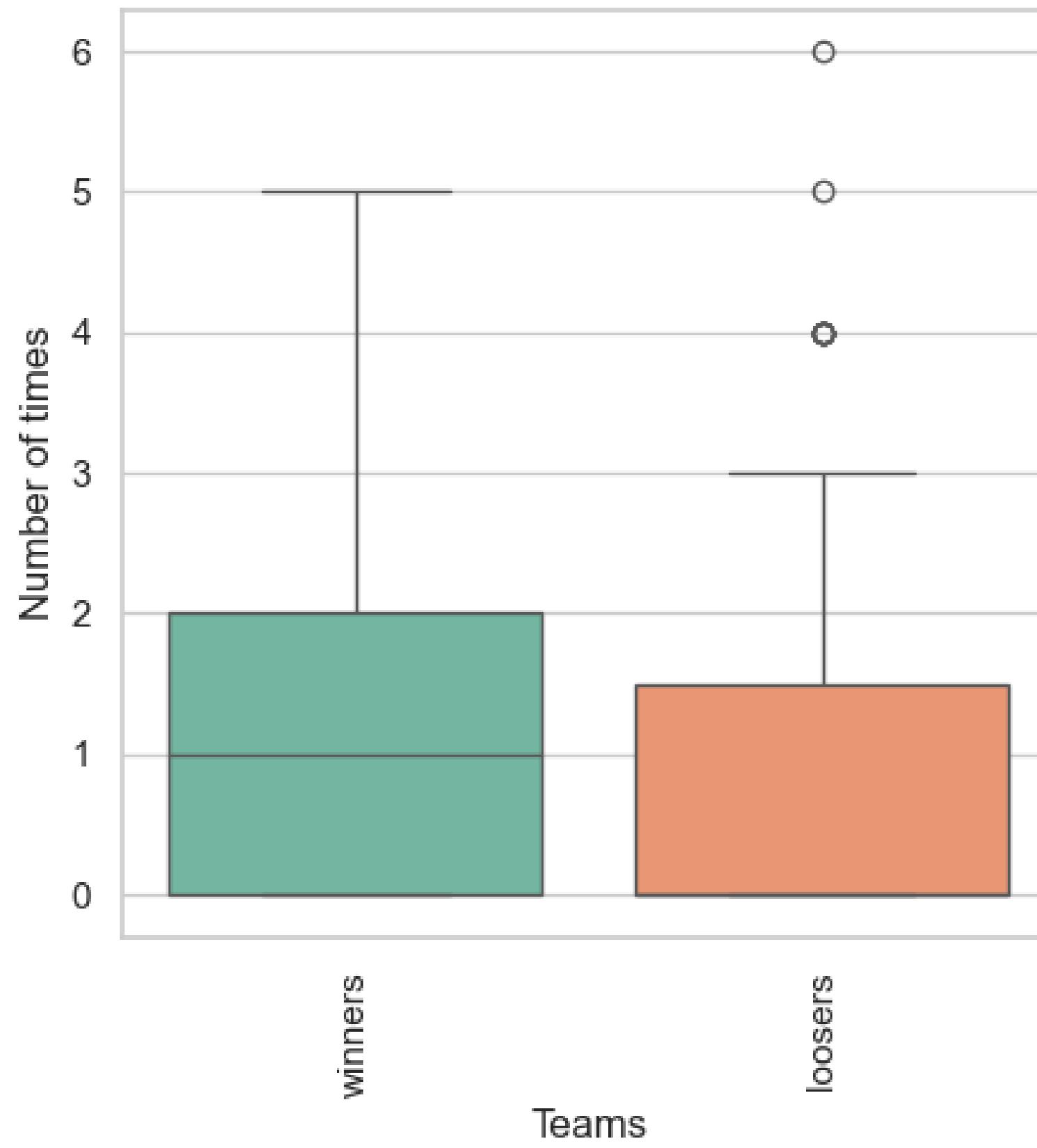
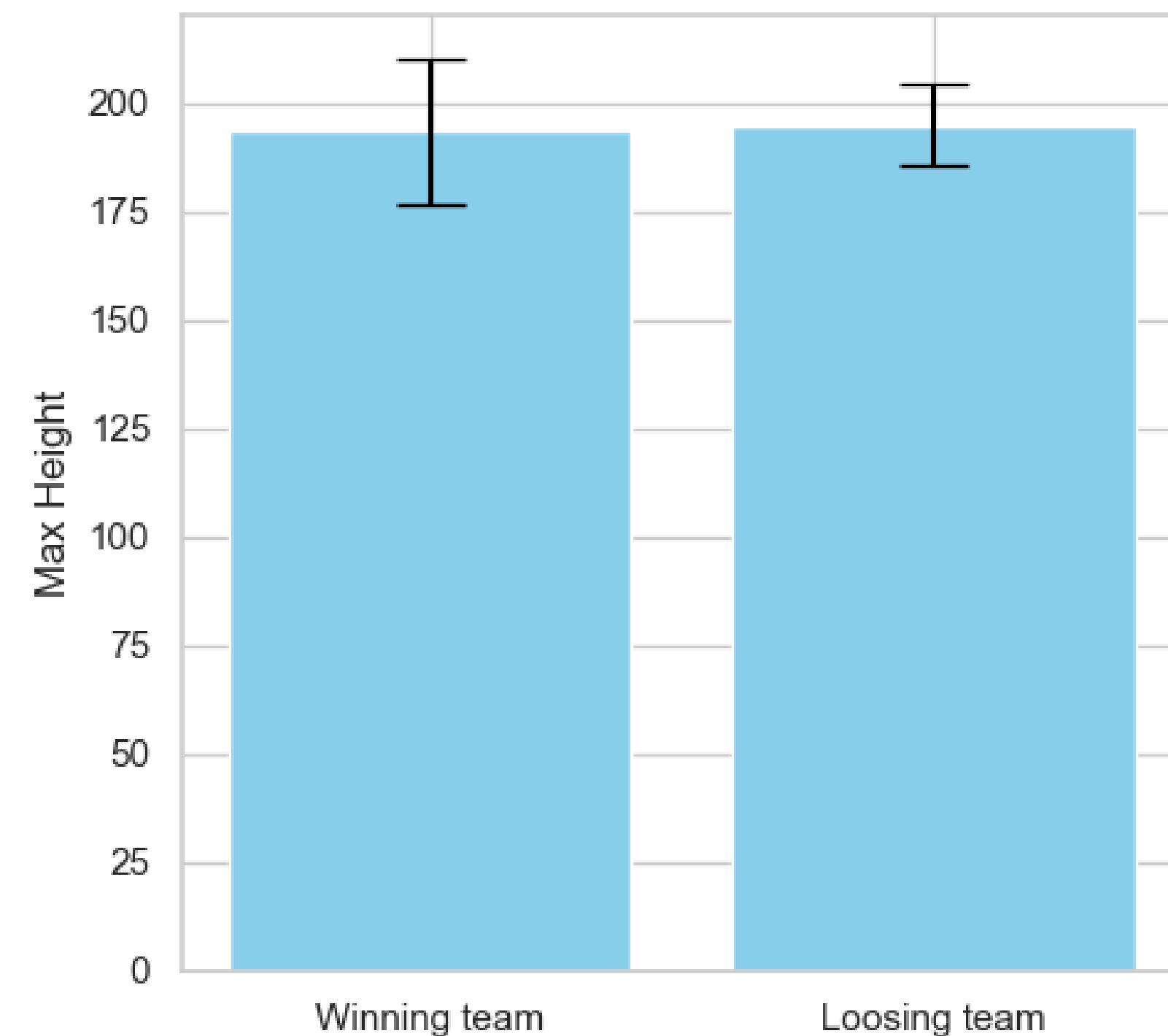


Fig: Distribution of left-footed players across winning and loosing

### Average of max height of winning teams vs loosing teams



### Max height of winning teams vs loosing teams

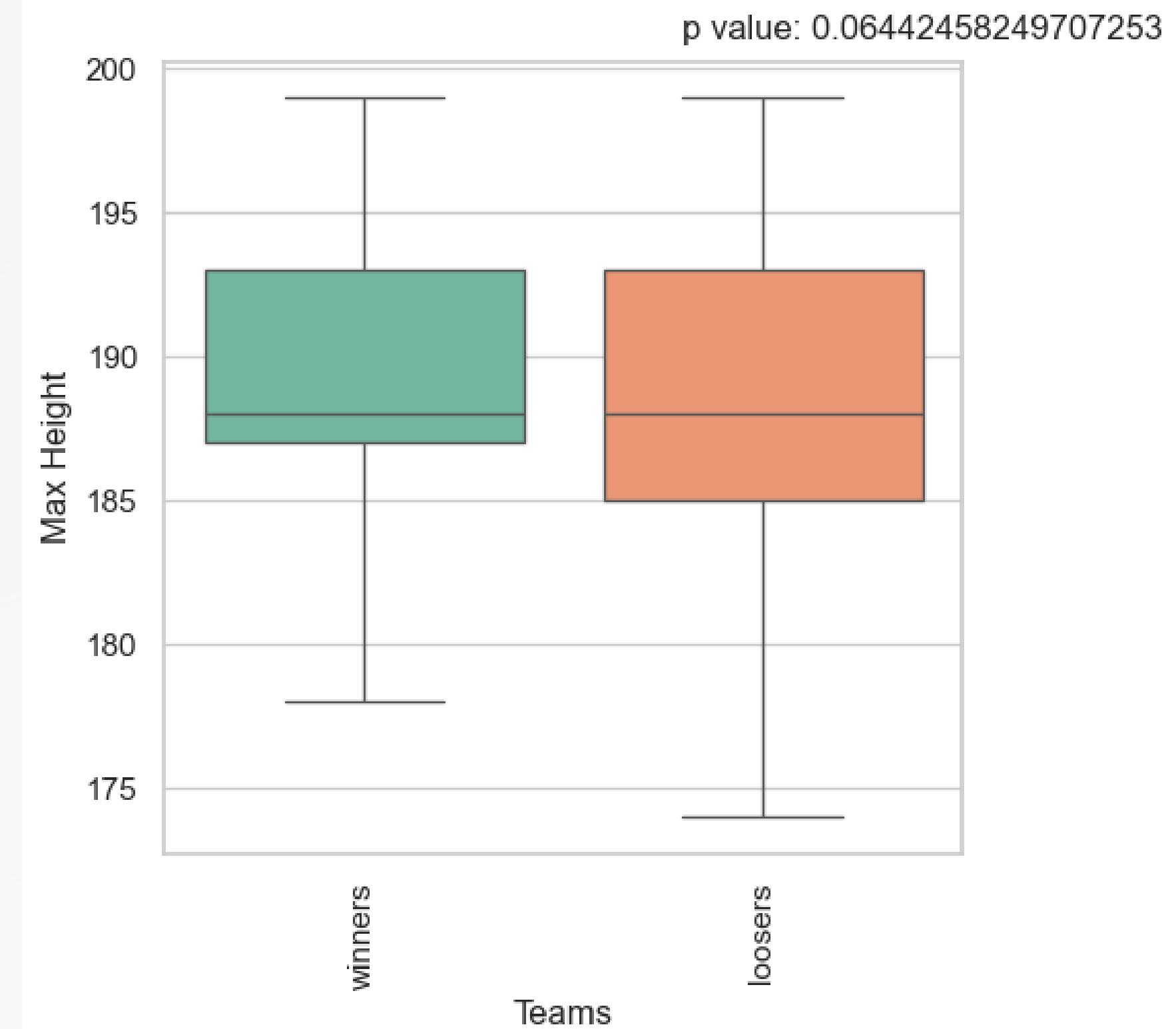
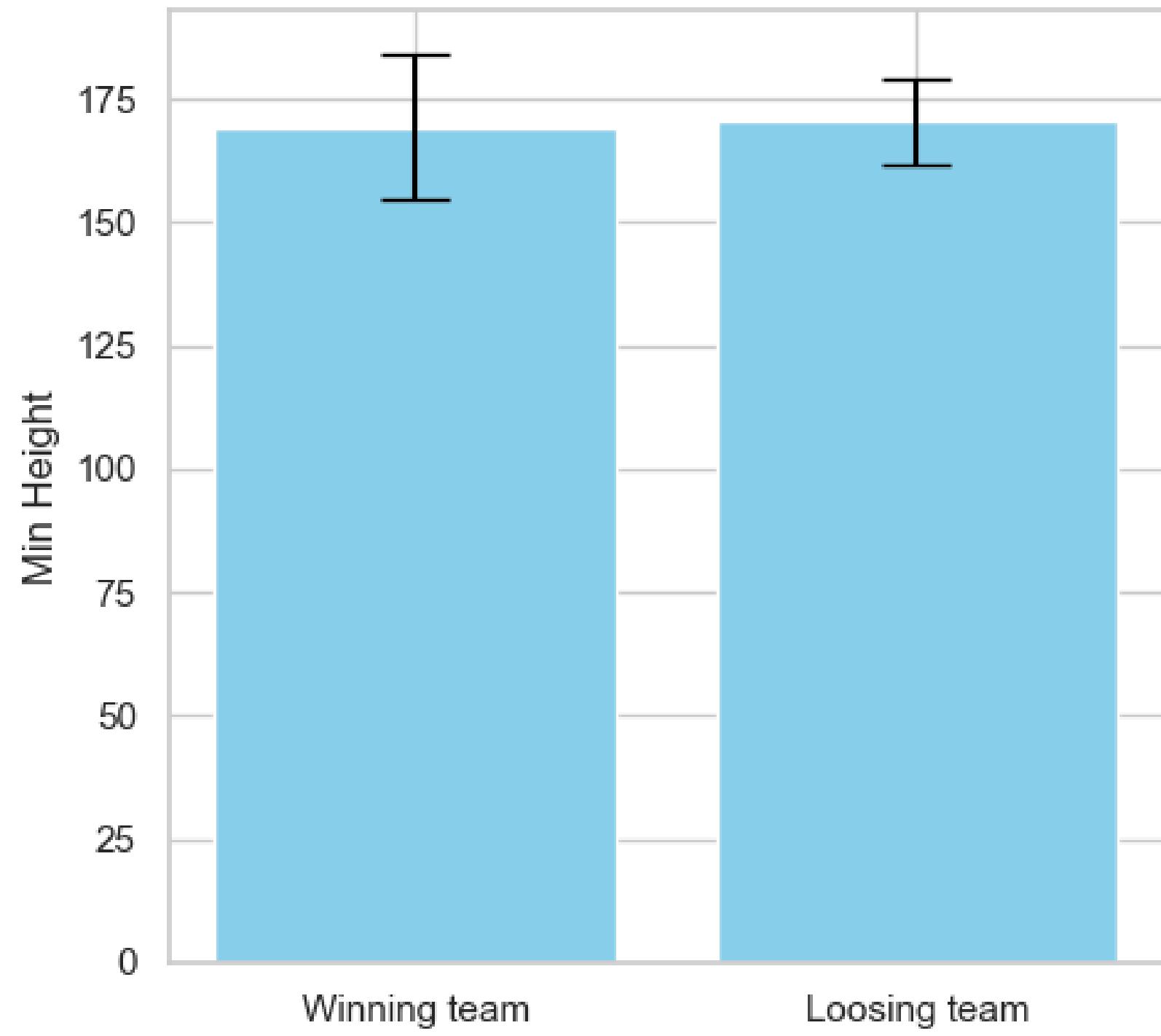


Fig: Footedness of the players in the tournament (England)

### Average of min. height of winning teams vs loosing teams



### Min height of winning teams vs loosing teams

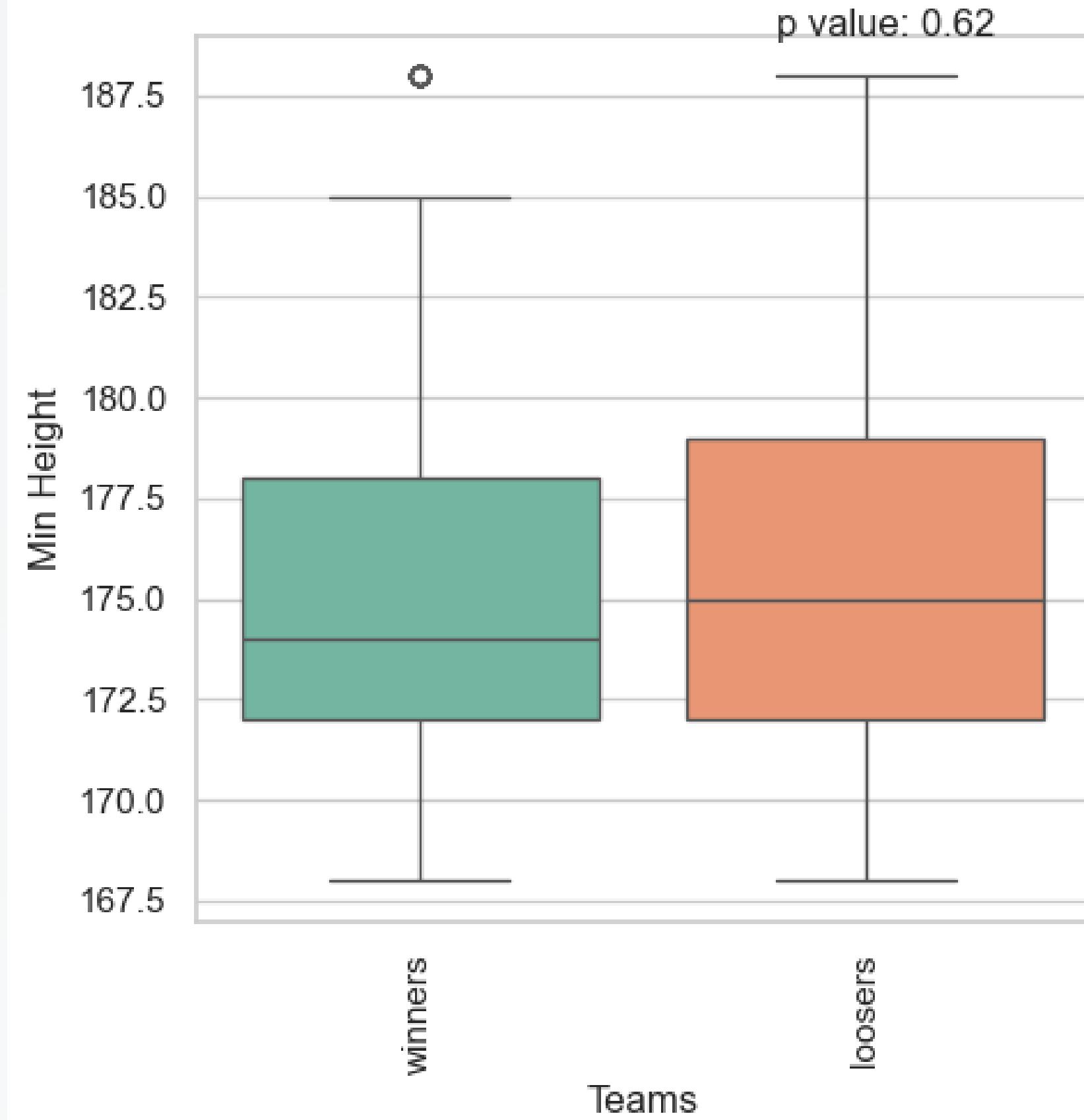
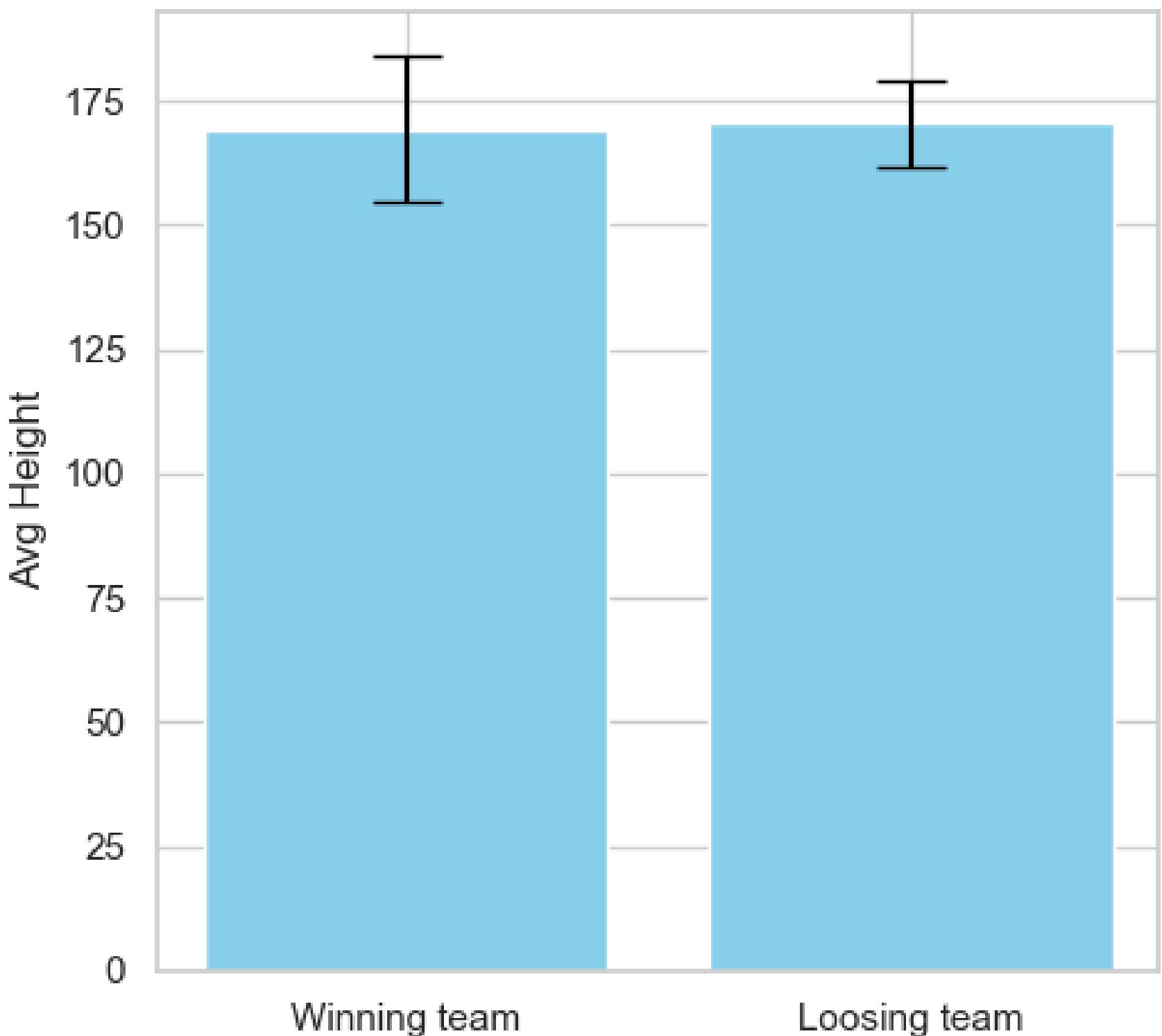


Fig: Footedness of the players in the tournament (England)

## Average of height of winning teams vs loosing teams



## Avg height of winning teams vs loosing teams

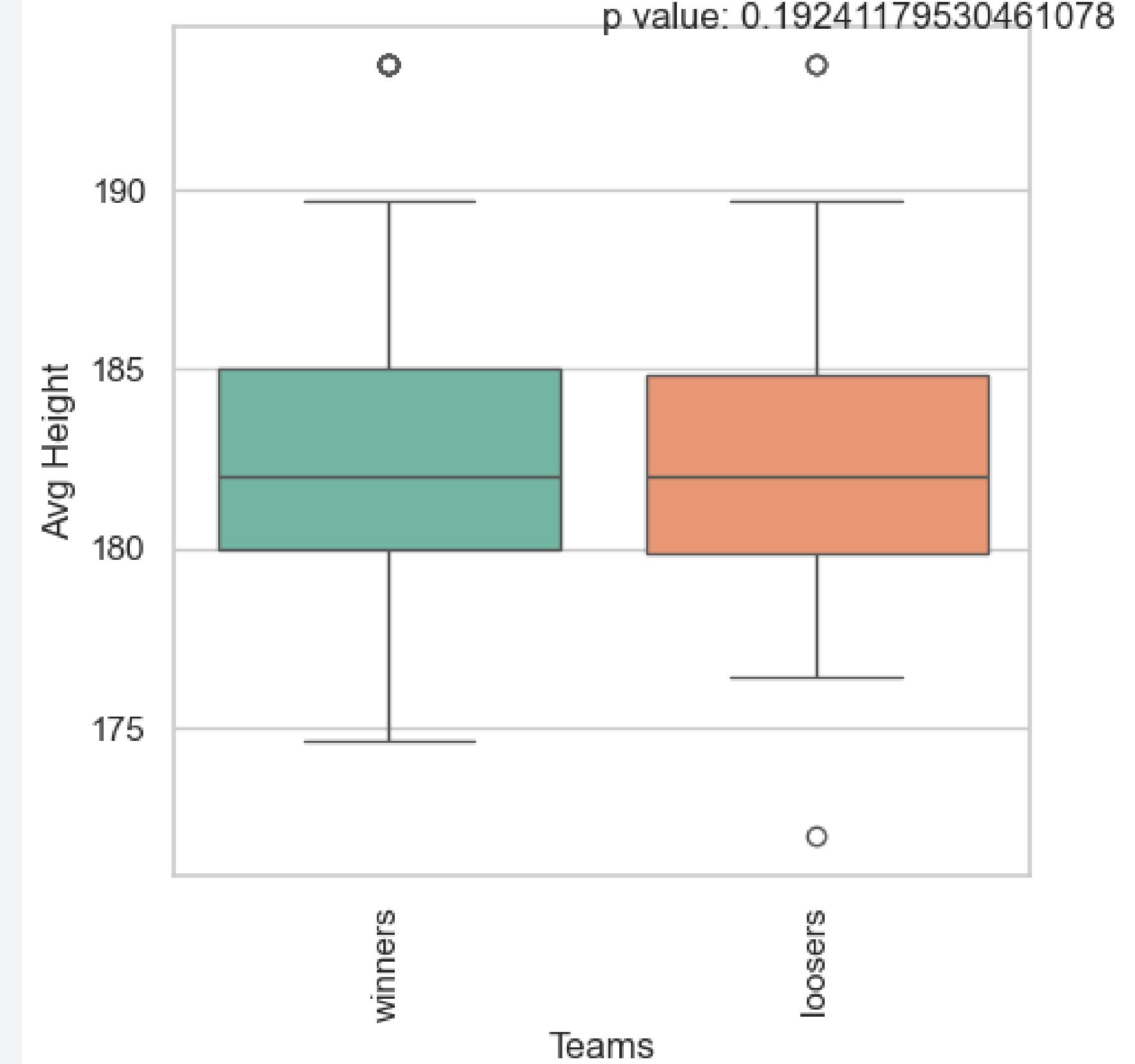
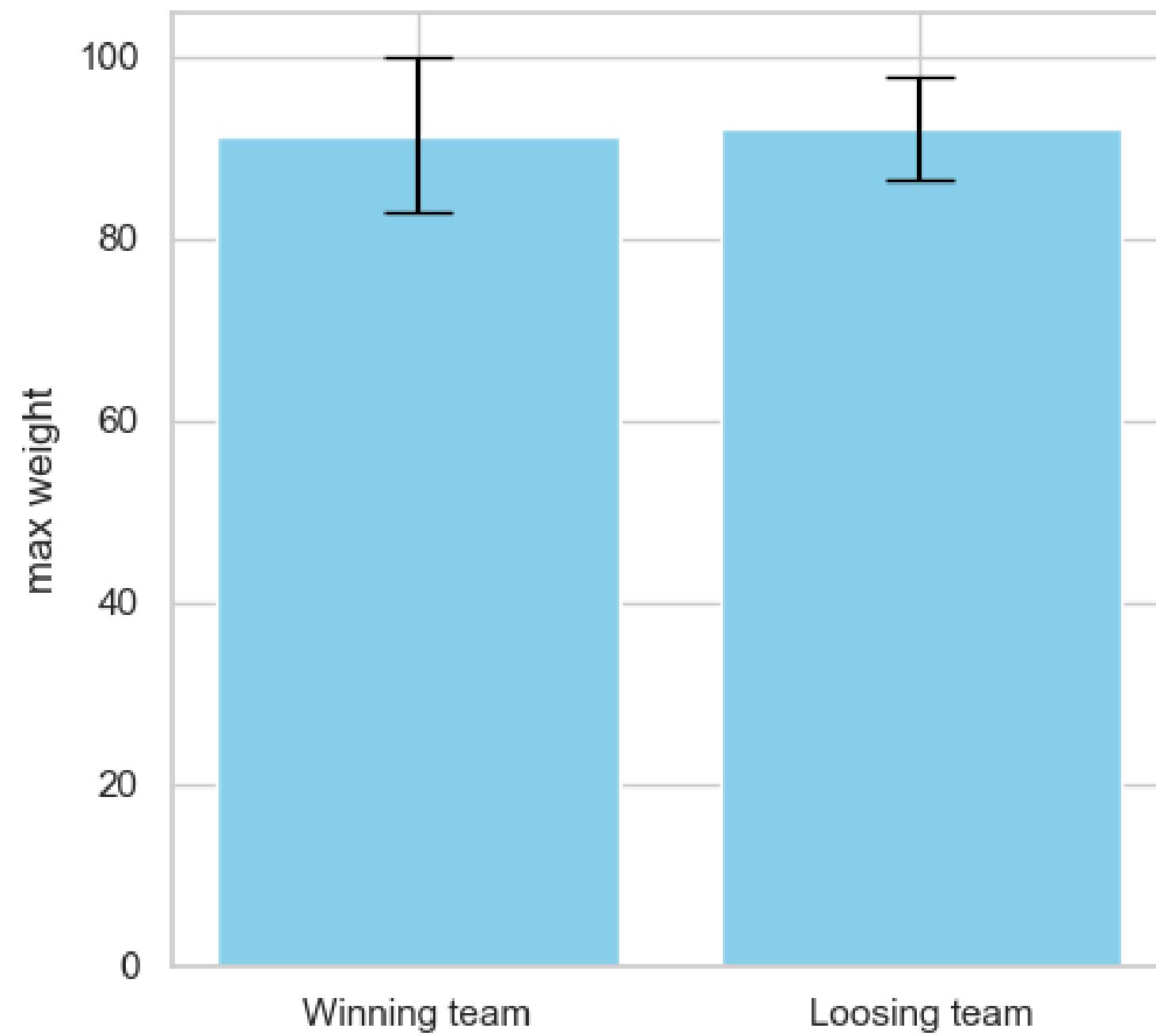


Fig: Footedness of the players in the tournament (England)

### Average of the max weights of winning teams vs loosing teams



### Max weight of winning teams vs loosing teams

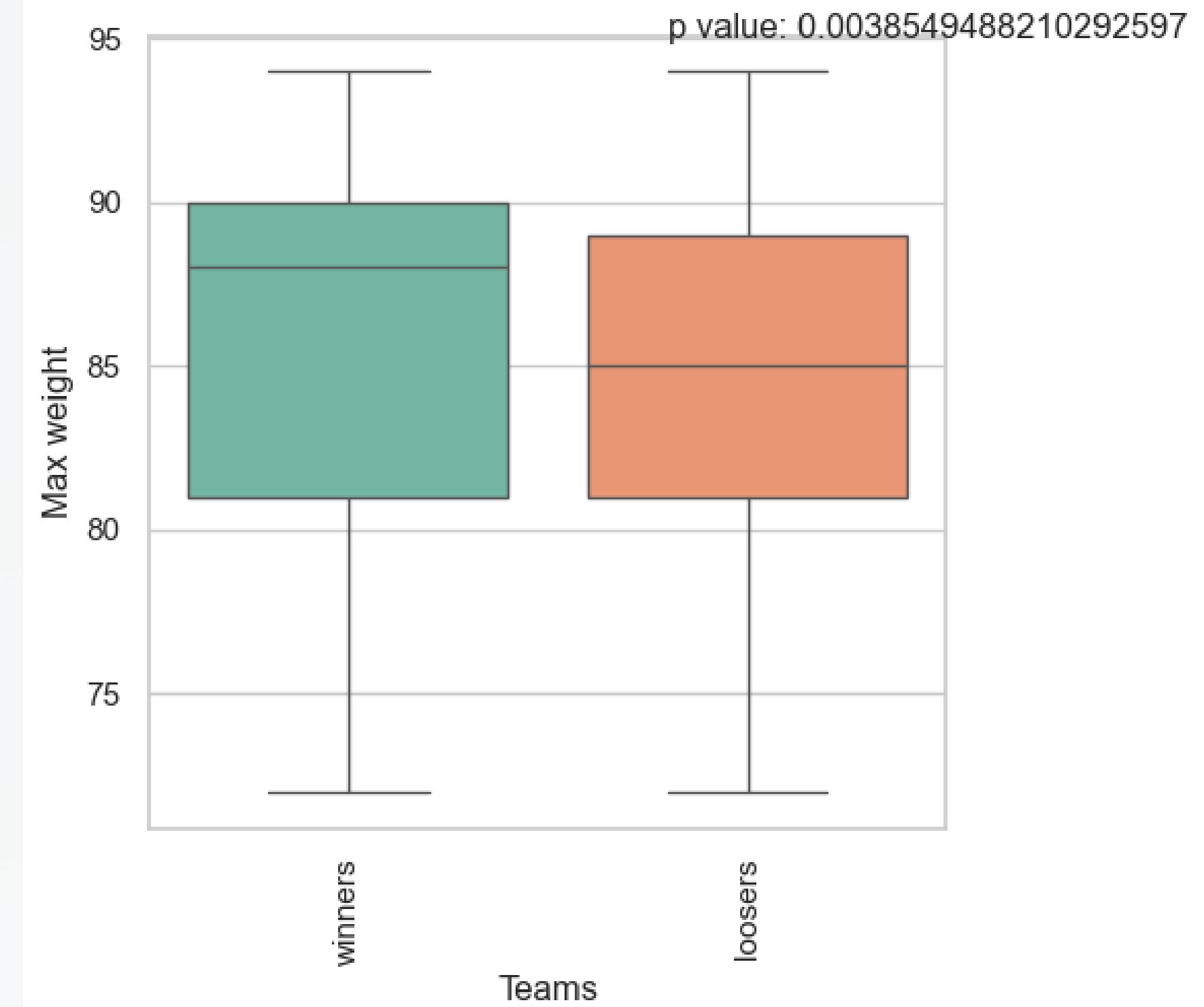


Fig: Footedness of the players in the tournament (England)

## Min weight of winning teams vs loosing teams

### Average of the min weights of winning teams vs loosing teams

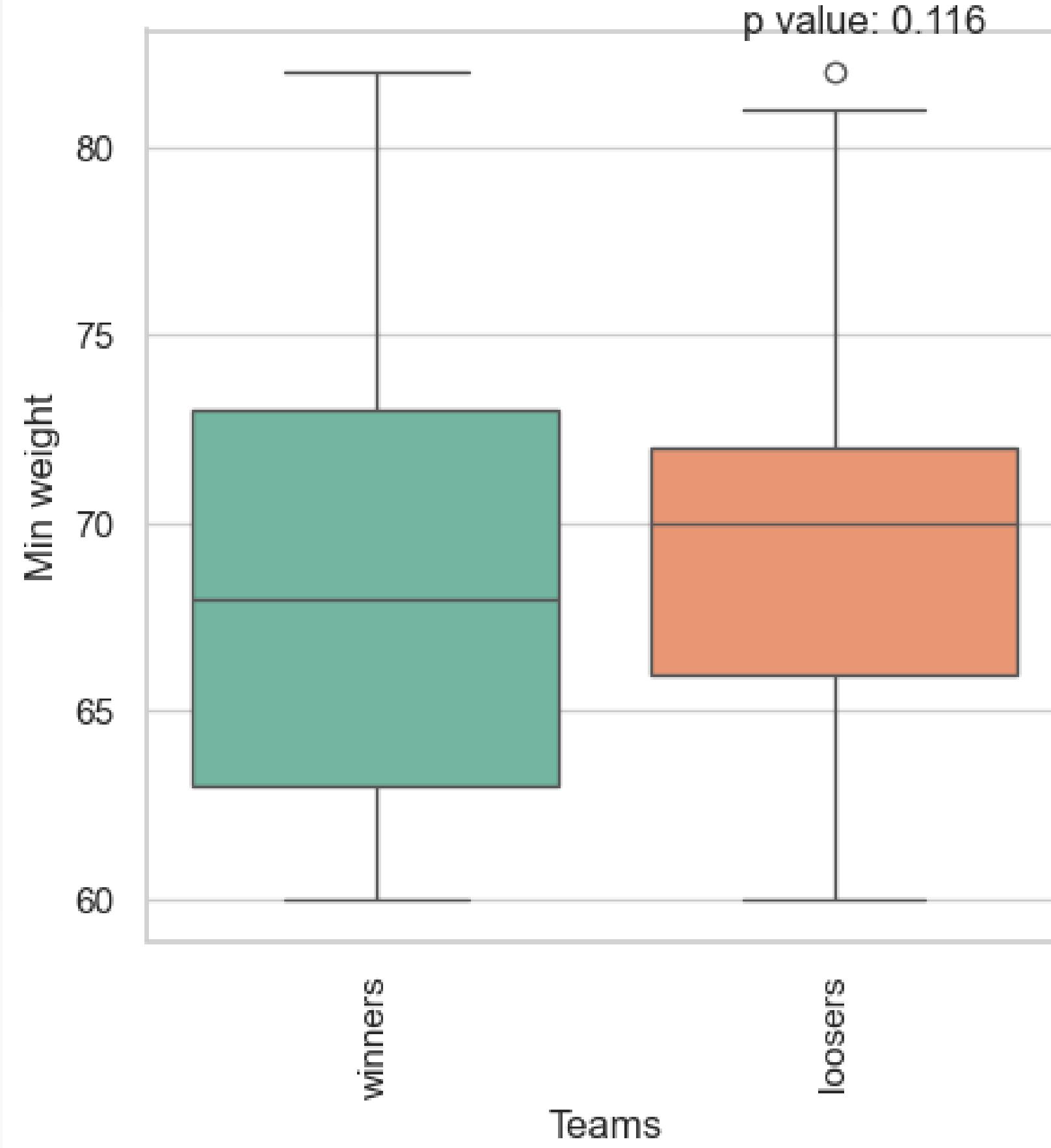
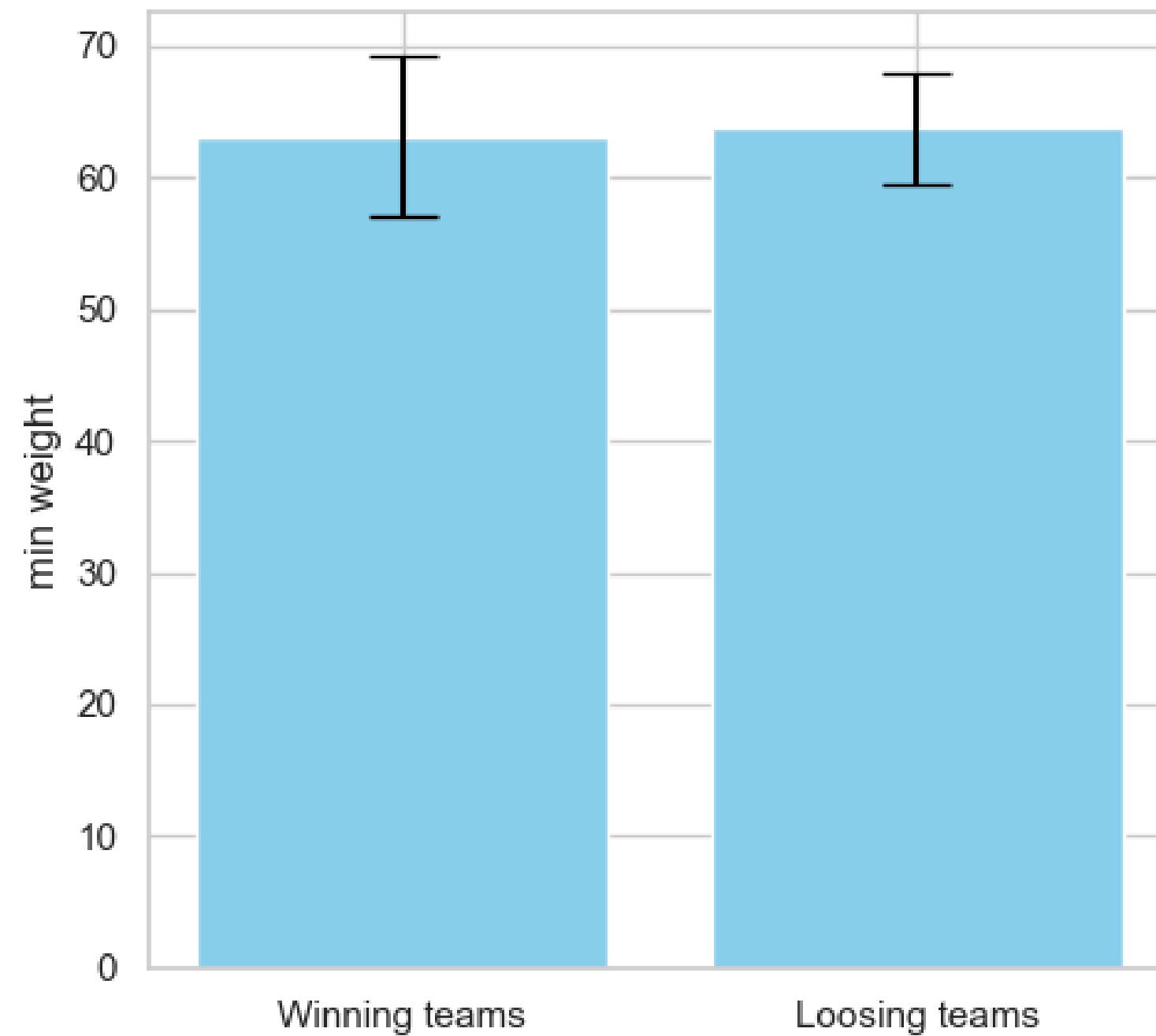
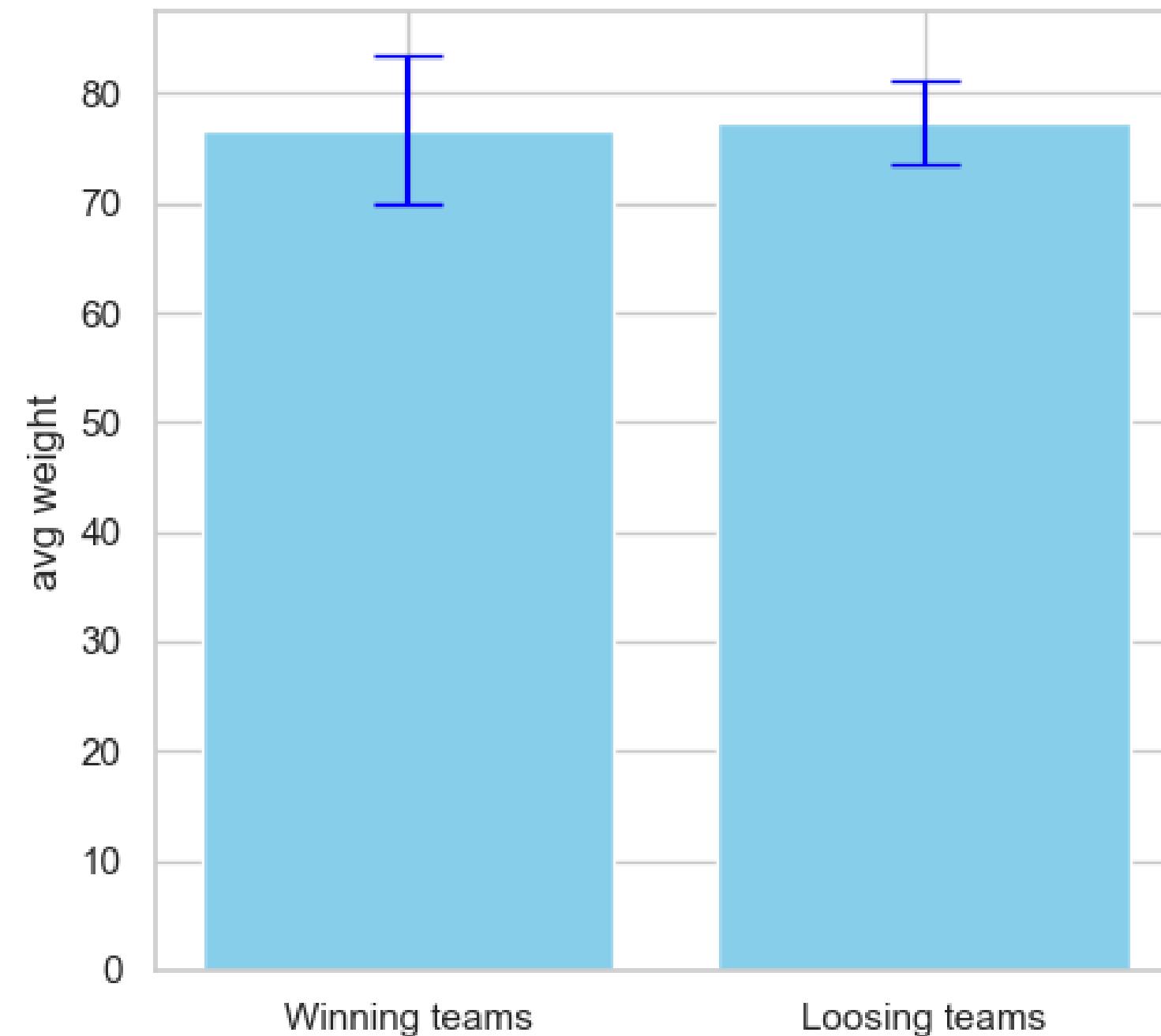


Fig: Footedness of the players in the tournament (England)

### Average of the avg weights of winning teams vs loosing teams



### Avg weight of winning teams vs loosing teams

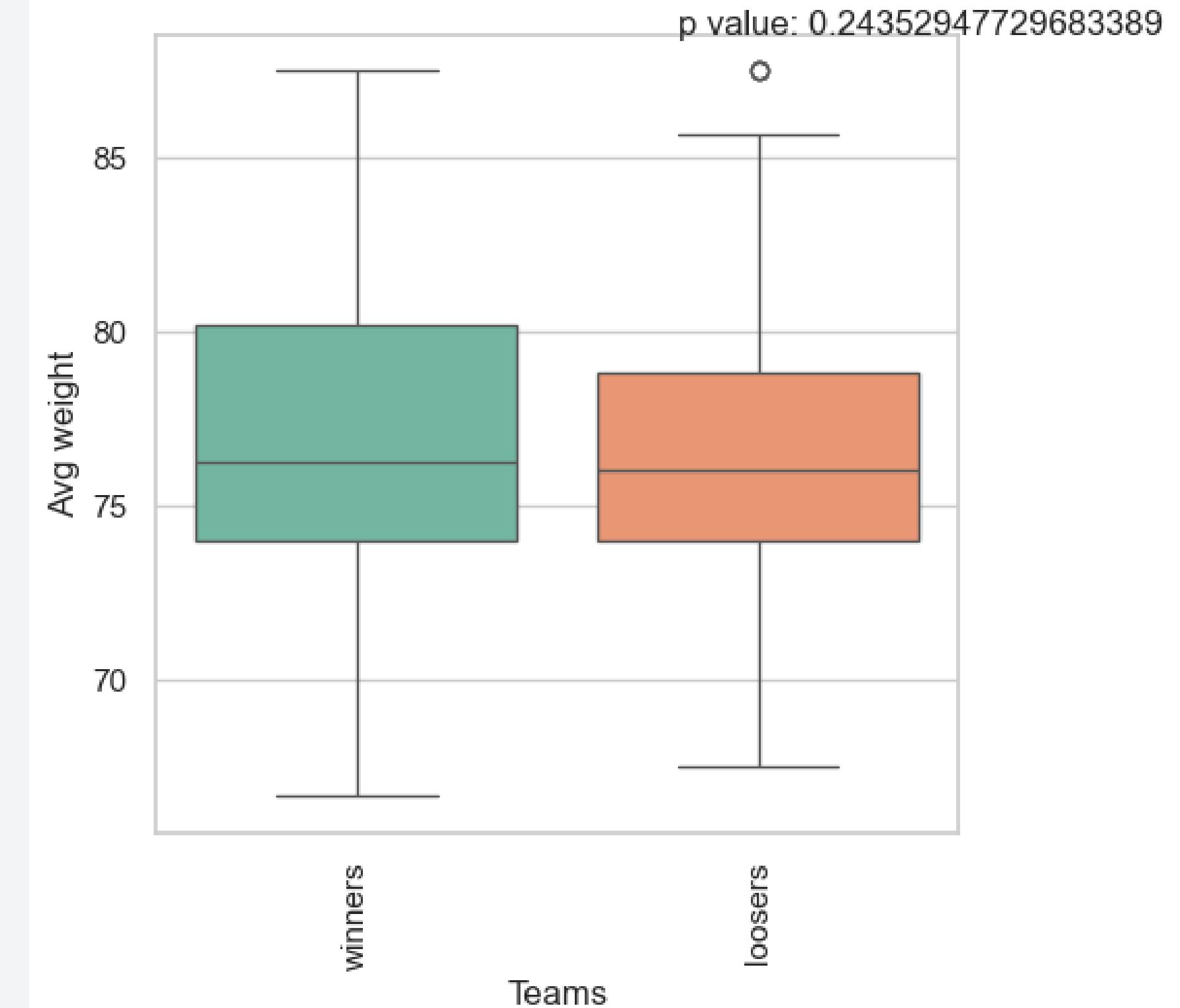


Fig: Footedness of the players in the tournament (England)

# CONCLUSION & DISCUSSION



We were unable to disprove the null hypothesis that the number of left-footed players does not impact the outcome of the match!



We Couldn't find any significant data to say there was an association between players' height and the outcome of the game.



There was a significant difference between the winning teams' heaviest players and the losing teams' heaviest players. This might be an interesting topic to explore

# ASSUMPTIONS



- Benched players were considered as if they were substituted into the match at some point.



- Own goals were considered mistakes and not included in individual players' goals



# LIMITATIONS

- Time in pitch for the players was not considered.
- Matches/teams/players were analysed in respective tournaments, not as a whole.
- Spatio-temporal data was not utilized.



# NEXT UP

Things that will be **neat** to try out

Take the playing time of each player into consideration when analysing the data

**STEP 1**

Explore the other tournaments, and take all of them as one and analyse the data.

**STEP 2**

Use the Spatio-Temporal data to make a network model and analyse the resulting data

**STEP 3**

# THANK YOU FOR LISTENING

