

# Building the Perfect Soccer Player

---

*Yi Chen, Liz Ehrlich, Dora Pekeč, Joe Zuo*

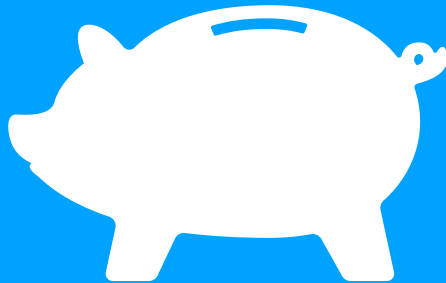
# Fast Facts

---



3.2 Billion

People watched at least a part of the  
2018 FIFA World Cup



€100 Million

To transfer Cristiano Ronaldo from Real Madrid  
to Juventus F. C.

A close-up photograph of Cristiano Ronaldo, a professional footballer, smiling and looking slightly to his left. He is wearing a white and black Juventus home jersey. The jersey features the Adidas logo on the right chest, the Italian national flag crest in the center, and the Juventus club crest on the left chest. The word 'Jeep' is printed in large white letters across the front. The background is a blurred stadium crowd.

# Cristiano Ronaldo

#7 / Forward

Juventus F. C.

5-time winner of Ballon d'Or

Born February 5, 1985

# The Dataset

---

- Data retrieved from [transfermarkt.com](https://www.transfermarkt.com) on Nov 27, 2018
- 500 top valued players so far in the 2018-2019 season
- European Leagues:

Majorly Premier League, La Liga, Bundesliga, Serie A, France Ligue 1

# The Variables

---

- **Market Value:**

The estimated monetary amount for a player to transfer

- **Position:**

The player's position on the field

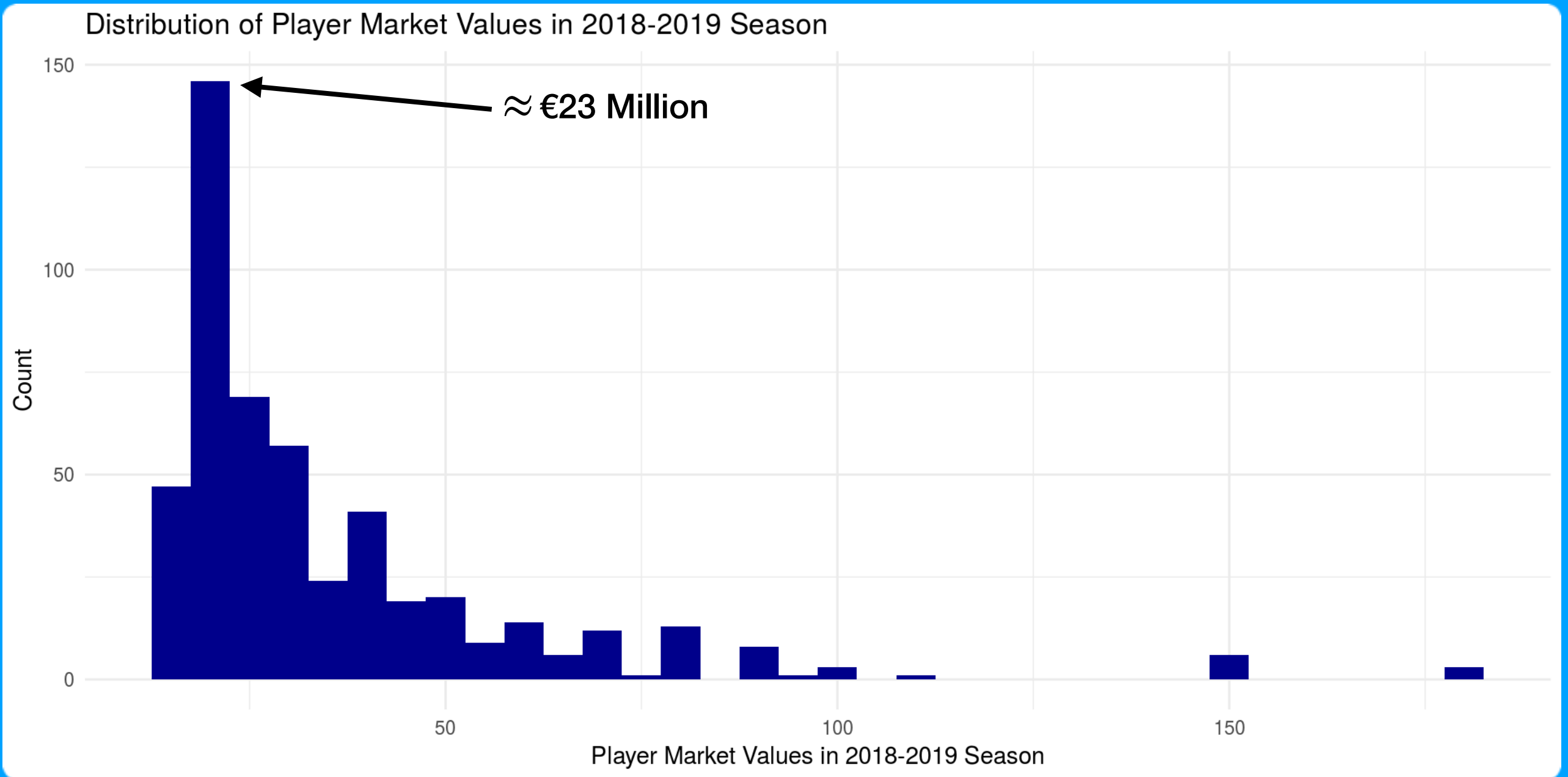
- **Goals and assists**

- **Yellow/red cards**

- **Other variables: age range, matches, substituted on/off**

# Data Analysis

## The Distribution

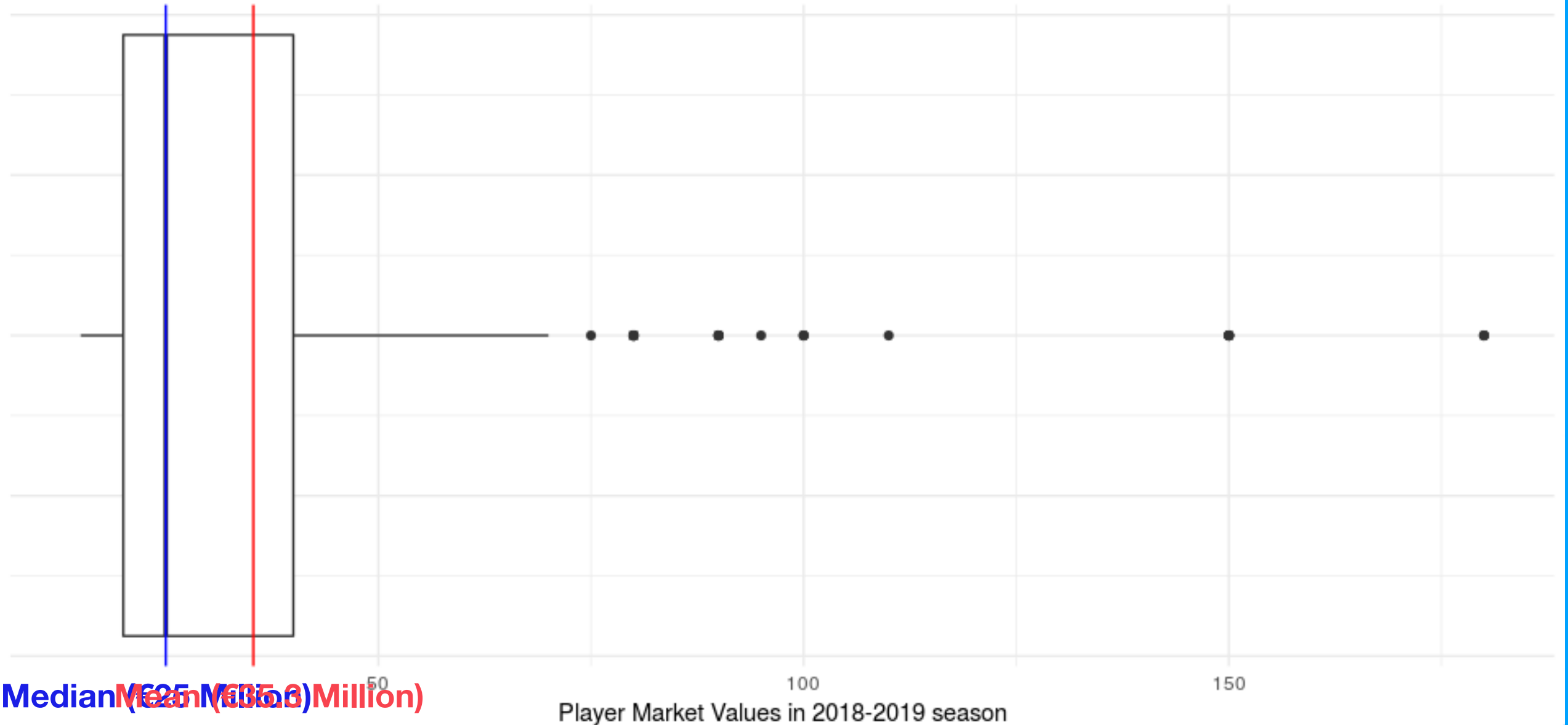




# Data Analysis

## The Distribution

Boxplot of Player Market Values in 2018-2019 Season



# Data Analysis

Do more expensive players contribute more?

---

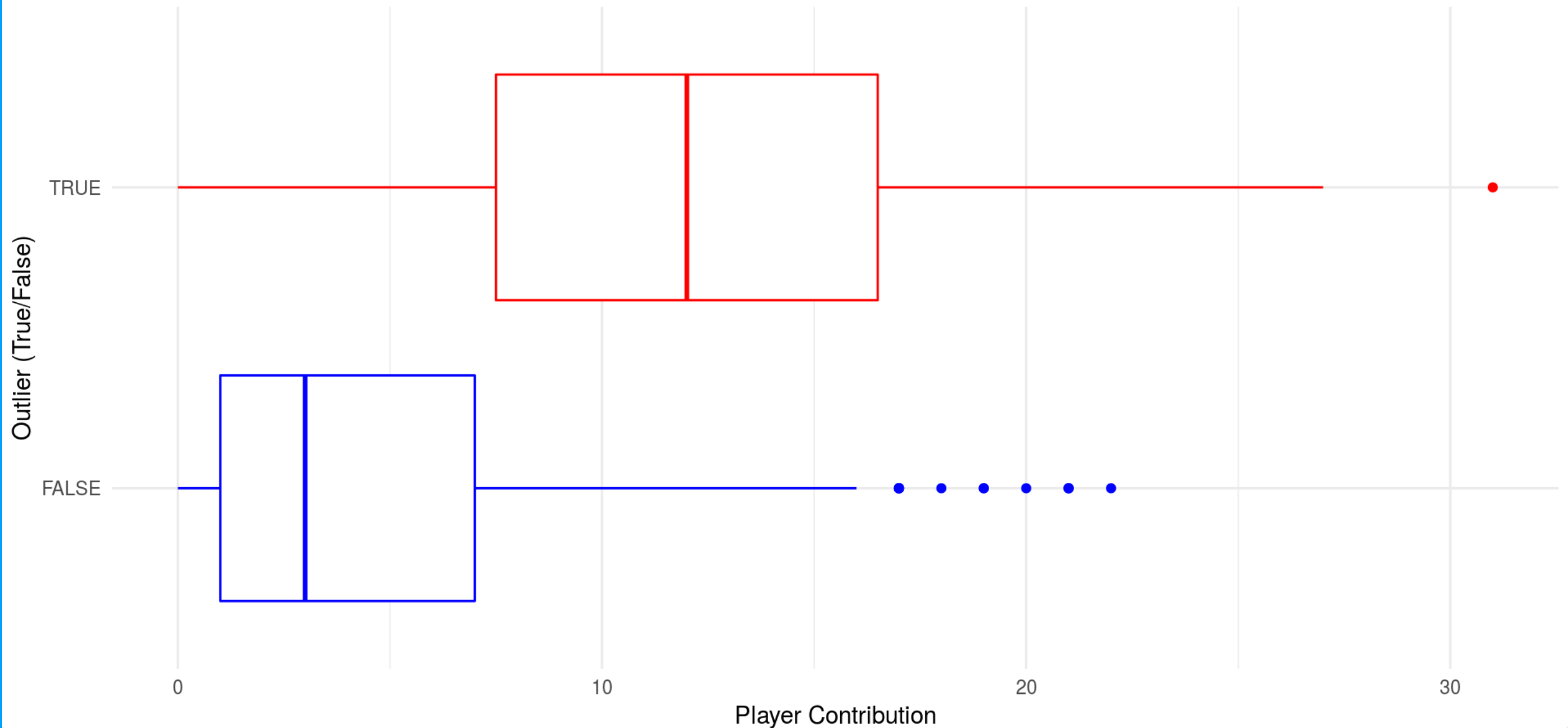
**Contribution = Goals + Assists**



# Data Analysis

Do more expensive players contribute more?

Boxplot of Player Contributions in 2018-2019 Season



# Data Analysis

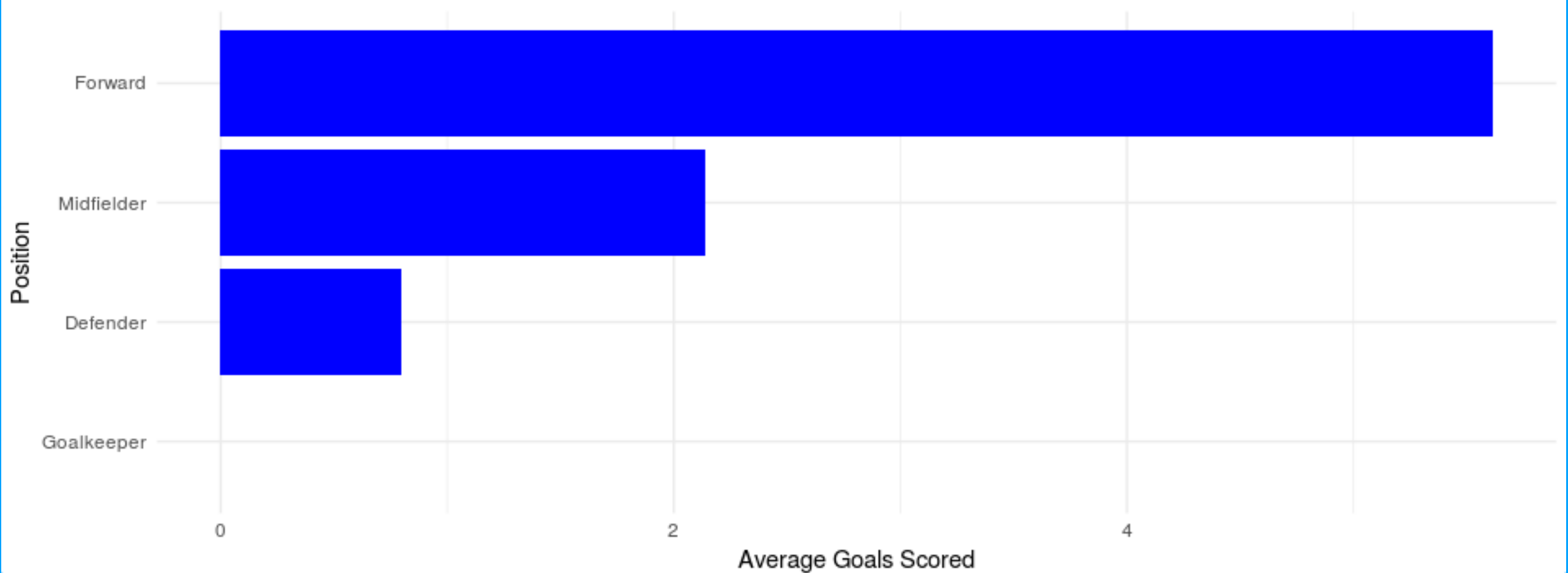
How do player positions affect goals and assists?



# Data Analysis

How do player positions affect goals and assists?

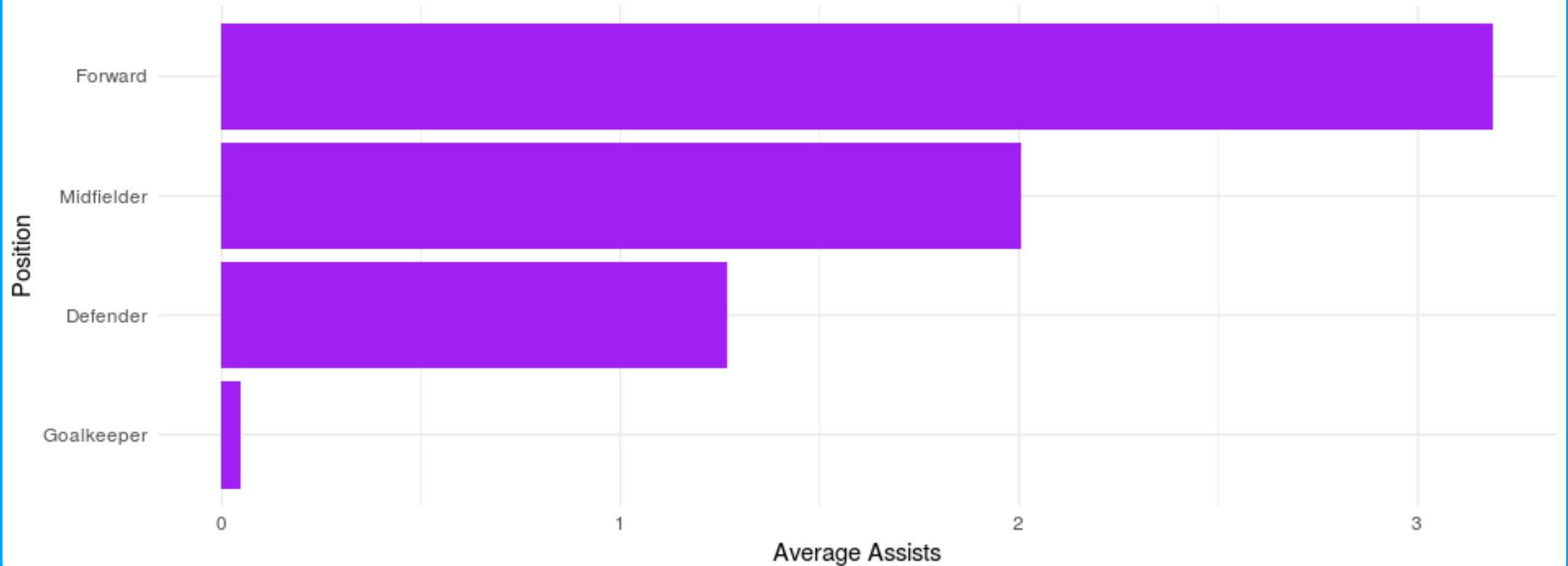
Average Goals Scored in Each Position in the 2018-2019 Season



# Data Analysis

How do player positions affect goals and assists?

Average Assists in Each Position in the 2018-2019 Season



# Data Analysis

How do player positions affect goals and assists?

---

**Ederson: Goalkeeper | 1 assist**

# What Makes a Valuable Player?

Building a linear model

```
## # A tibble: 21 x 5
```

##	term	estimate	std.error	statistic	p.value
##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	factor(position_new)Defender	12.6	15.4	0.818	0.414
## 2	factor(position_new)Forward	-0.965	15.7	-0.0616	0.951
## 3	factor(position_new)Goalkeeper	17.2	16.1	1.07	0.287
## 4	factor(position_new)Midfielder	18.1	15.4	1.17	0.242
## 5	age	0.112	0.734	0.153	0.879
## 6	matches	0.579	0.219	2.64	0.00852
## 7	goals	2.04	1.66	1.23	0.220
## 8	own_goals	-0.367	5.66	-0.0649	0.948
## 9	assists	0.789	1.12	0.707	0.480
## 10	yellow_cards	-0.515	0.599	-0.859	0.391
## #	... with 11 more rows				

# What Makes a Valuable Player?

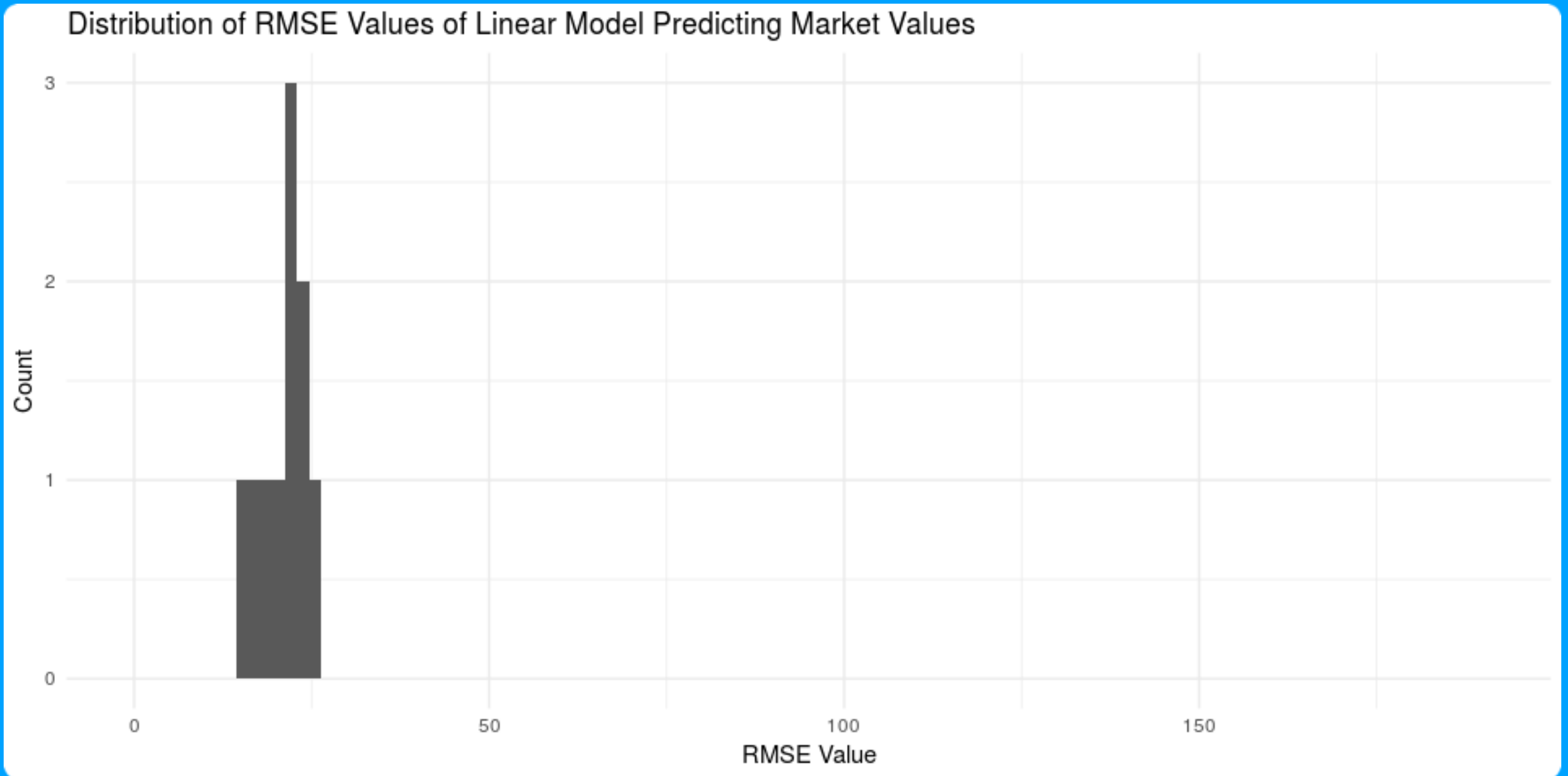
Building a linear model

```
## # A tibble: 15 x 5
##   term                                estimate std.error statistic p.value
##   <chr>                                <dbl>    <dbl>    <dbl>    <dbl>
## 1 factor(position_new)Defender         15.9      5.35      2.98 3.04e-3
## 2 factor(position_new)Forward        -0.607     5.57     -0.109 9.13e-1
## 3 factor(position_new)Goalkeeper      22.7      7.05      3.22 1.36e-3
## 4 factor(position_new)Midfielder      19.6      5.43      3.61 3.32e-4
## 5 matches                             0.406     0.179     2.26 2.42e-2
## 6 goals                               2.02      1.64      1.23 2.21e-1
## 7 assists                             0.885     1.10      0.802 4.23e-1
## 8 factor(age_range)21-25               3.66      4.49      0.815 4.15e-1
## 9 factor(age_range)26-30              9.07      4.53      2.00 4.57e-2
## 10 factor(age_range)30 and above      -0.922     5.71     -0.162 8.72e-1
## 11 factor(position_new)Forward:goals   0.937      1.68      0.557 5.78e-1
## 12 factor(position_new)Midfielder:go... -1.45      1.85     -0.783 4.34e-1
## 13 factor(position_new)Forward:assis... 2.74      1.28      2.14 3.25e-2
## 14 factor(position_new)Goalkeeper:as... 23.4      22.1      1.06 2.90e-1
## 15 factor(position_new)Midfielder:as... -0.366     1.52     -0.241 8.09e-1
```



# What Makes a Valuable Player?

How good is our model?



# What Makes a Valuable Player?

How good is our model?

---

```
## # A tibble: 1 x 5
##       min    max  mean median    sd
##   <dbl> <dbl> <dbl>  <dbl> <dbl>
## 1  15.5  25.3  21.0   21.9  3.12
```

# What Makes a Valuable Player?

How good is our model?

---



We have a good model!

# Conclusion

What did we do?

---

- Visualized distribution of market values
- Identified outliers and explained them using data
- Explored goals and assists by position
- Created linear model
- Validated our model

# Conclusion

What did we find?

---

- Players' baseline market values
- Variables contributing to higher market values
- Interaction effects between variables

# Conclusion

How can we improve?

---

- Verify our sources
- Introduce historical data
- Use larger sample

**Thank you!**

---