

DATA ANALYSIS PROJCT

TEAM MEMBERS:
HOSSEIN FAGHIHI
AMIRHOSSEIN RANJBAR

DATA DESCRIPTION

The data is from the [Tennis Database]. The data contains information about the players, their matches, and the matches they played in. The data was in the form of a zip file. The zip file contains 31 zipfiles that contains parquet files for each table in the database. Each parquet file contains the data for a single table for just one day.

1. HOW MANY TENNIS PLAYERS ARE INCLUDED IN THE DATASET?

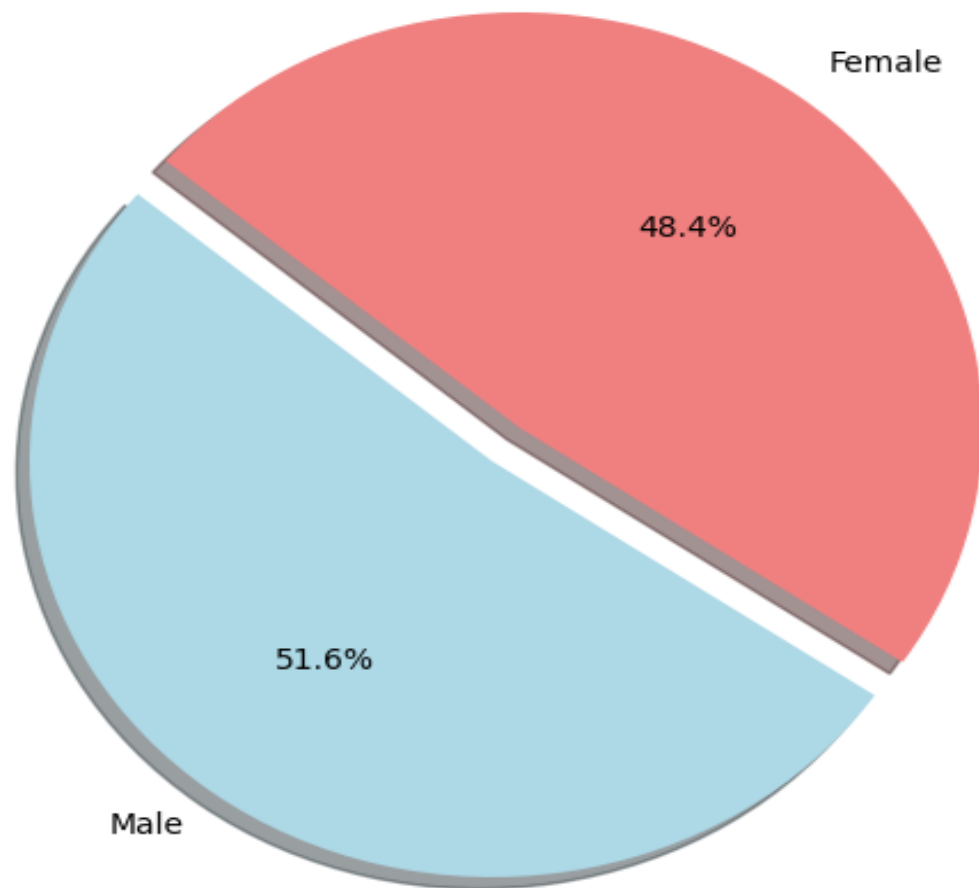
The number of unique tennis players in the dataset is: 2353

The average height of female players is: 1136

The average height of male players is: 1211

6 players gender is unknown

Distribution of Male and Female Tennis Players
Total Players: 2353



Gender Distribution

- Male
- Female

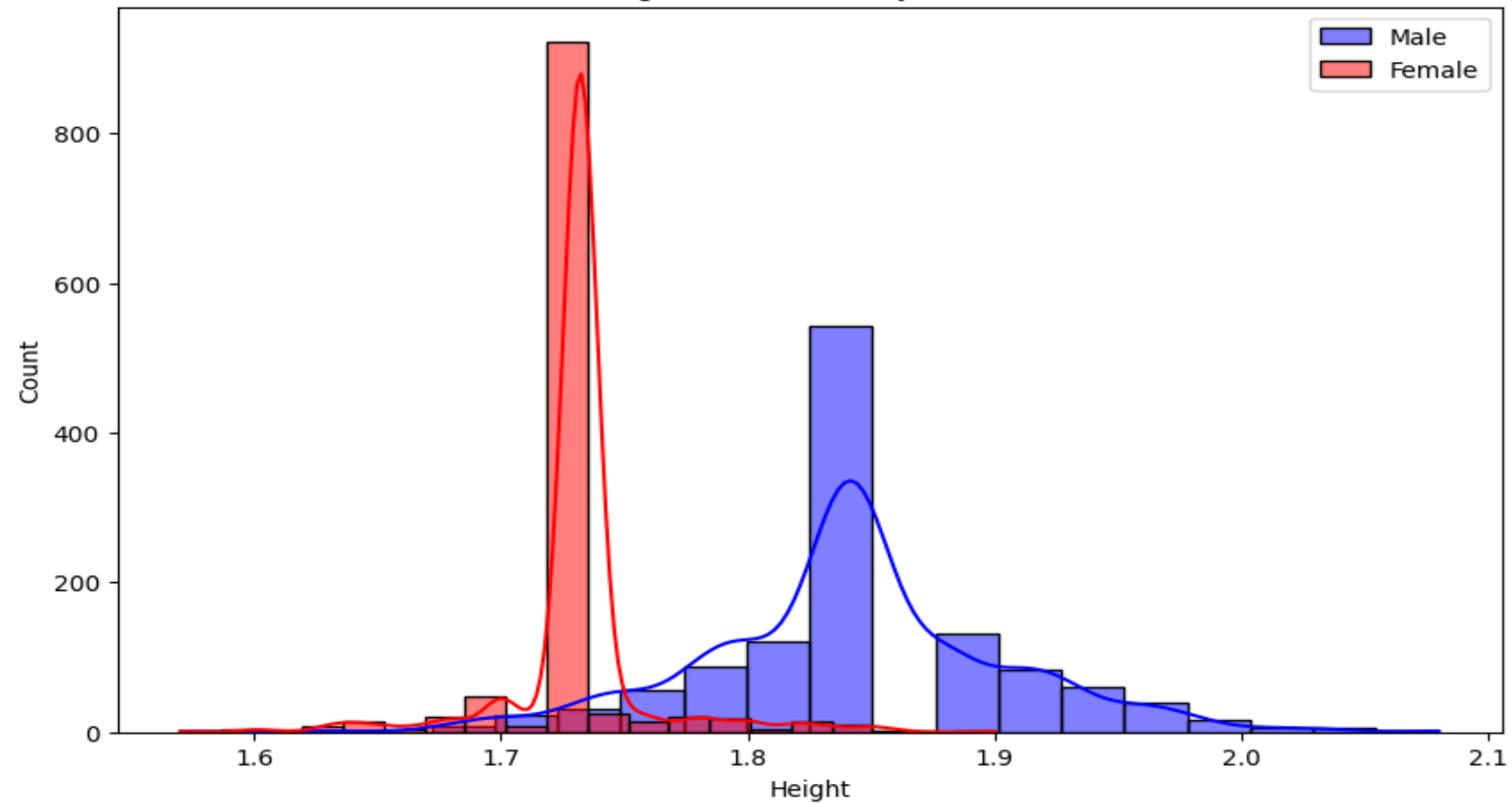
2. WHAT IS THE AVERAGE HEIGHT OF THE PLAYERS?

The average height of female players is 1.7321576763485476

The average height of male players is 1.8423621227887617

The average height of players is 1.7891325324733314

Height Distribution by Gender



3.1. WHICH PLAYER HAS THE HIGHEST NUMBER OF WINS?



Darder L



Juszcza P

2 male players with same number of wins:

Male player with the highest number of wins: Juszcza P. with 13 wins

Male player with the highest number of wins: Darder L. with 13 wins

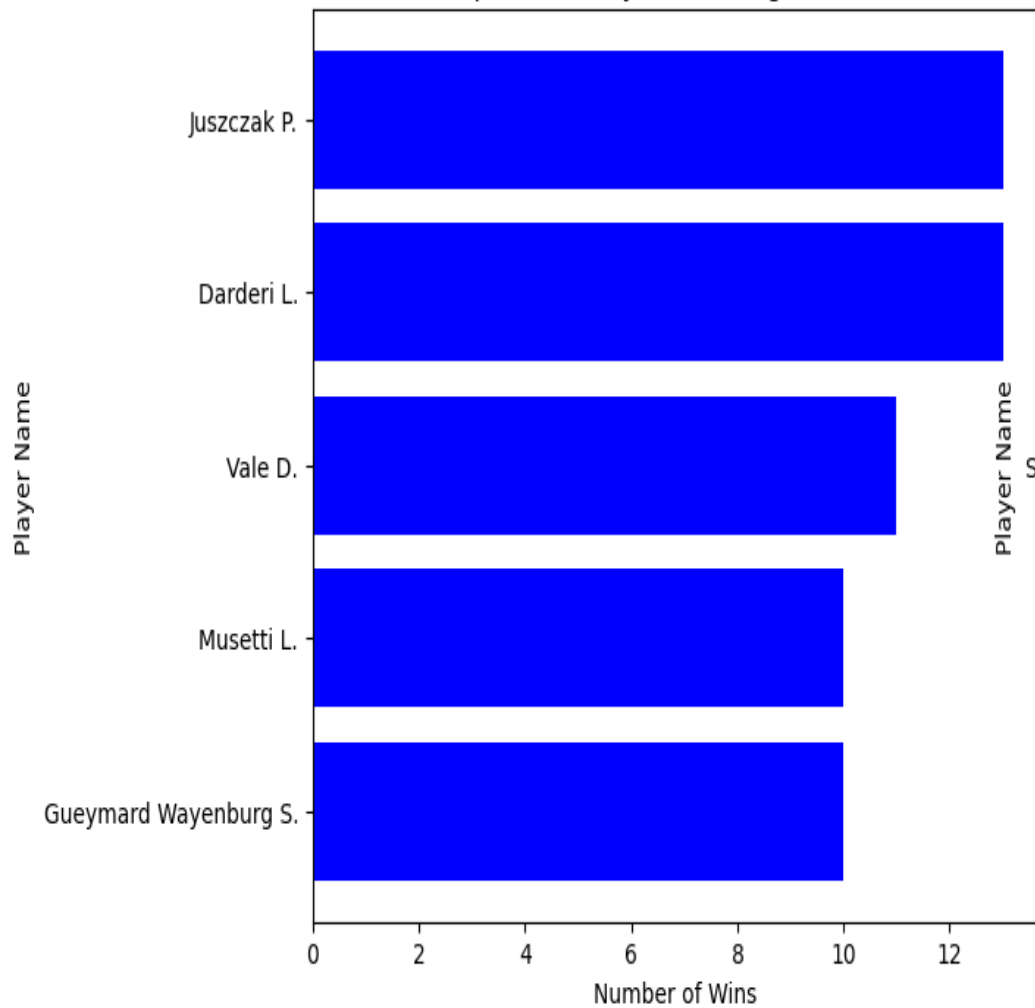
3.2. WHICH PLAYER HAS THE HIGHEST NUMBER OF WINS?

2 female players with same number of wins:

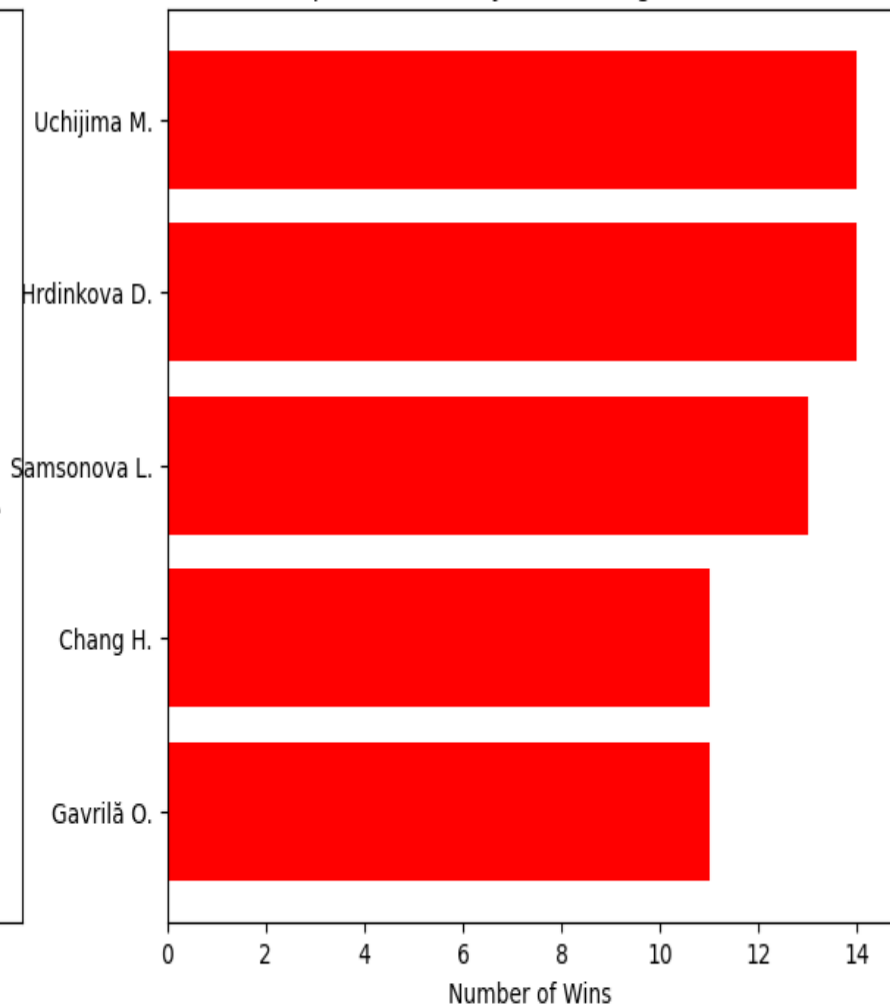
Female player with the highest number of wins: Uchijima M. with 14 wins

Female player with the highest number of wins: Hrdinkova D. with 14 wins

Top 5 Male Players with Highest Wins



Top 5 Female Players with Highest Wins



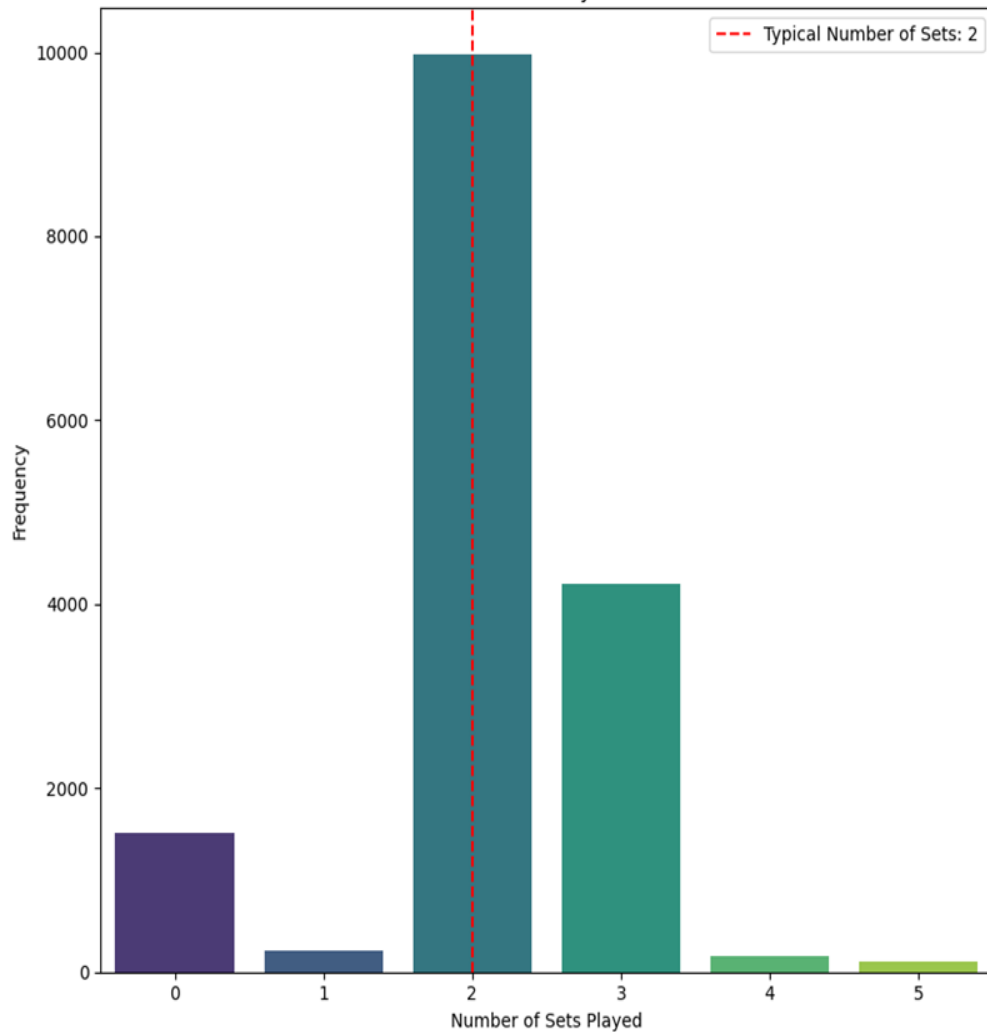
4. WHAT IS THE LONGEST MATCH RECORDED IN TERMS OF
DURATION?

The longest match_id was 12381286 and lasted 7.843333333333334 hours

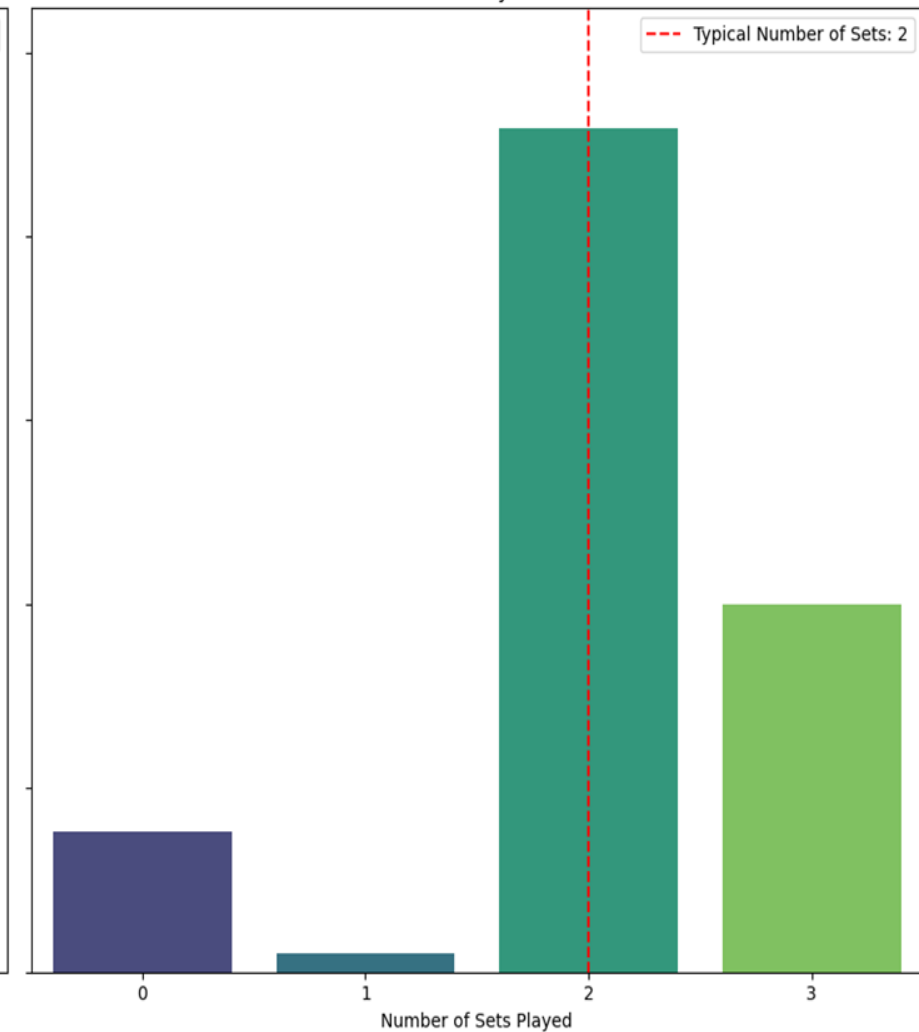
5. HOW MANY SETS ARE TYPICALLY PLAYED IN A TENNIS MATCH?

Typically, 2 sets are played in a male tennis match.
Typically, 2 sets are played in a female tennis match.

Distribution of Number of Sets Played in Male Tennis Matches



Distribution of Number of Sets Played in Female Tennis Matches

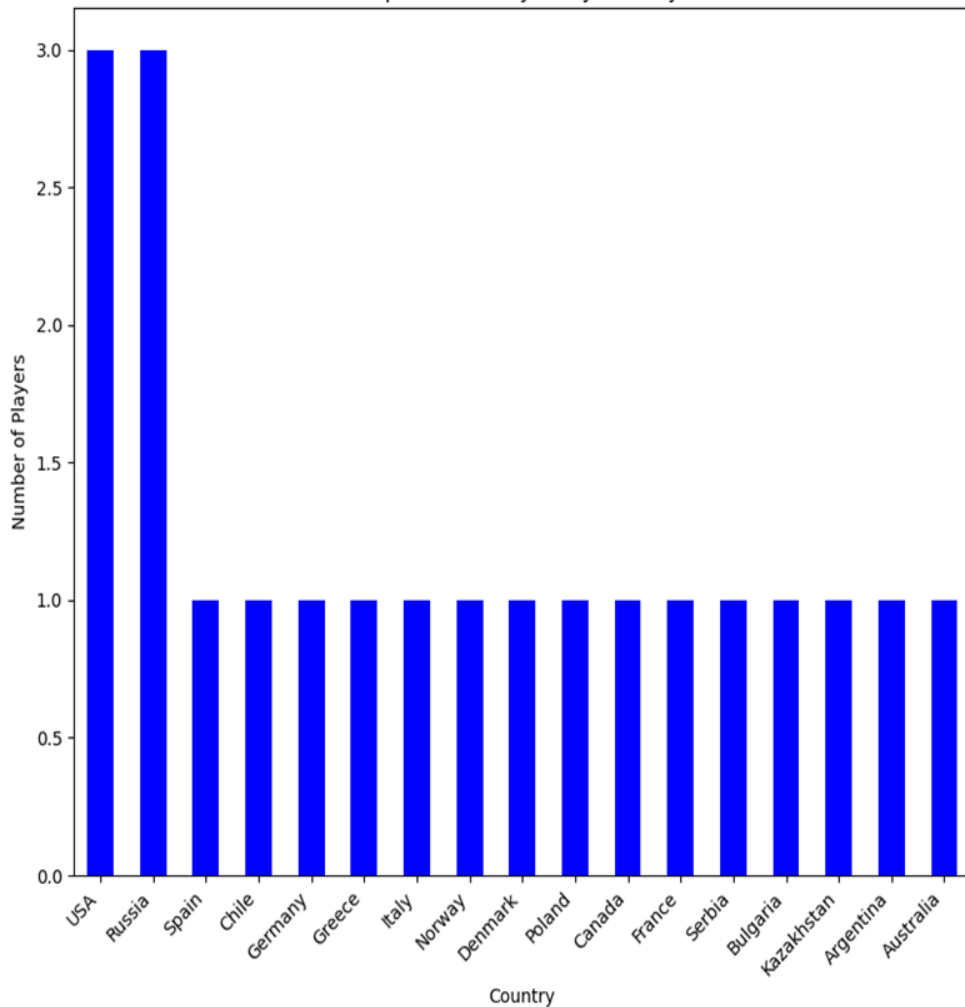


6. WHICH COUNTRY HAS PRODUCED THE MOST SUCCESSFUL TENNIS PLAYERS?

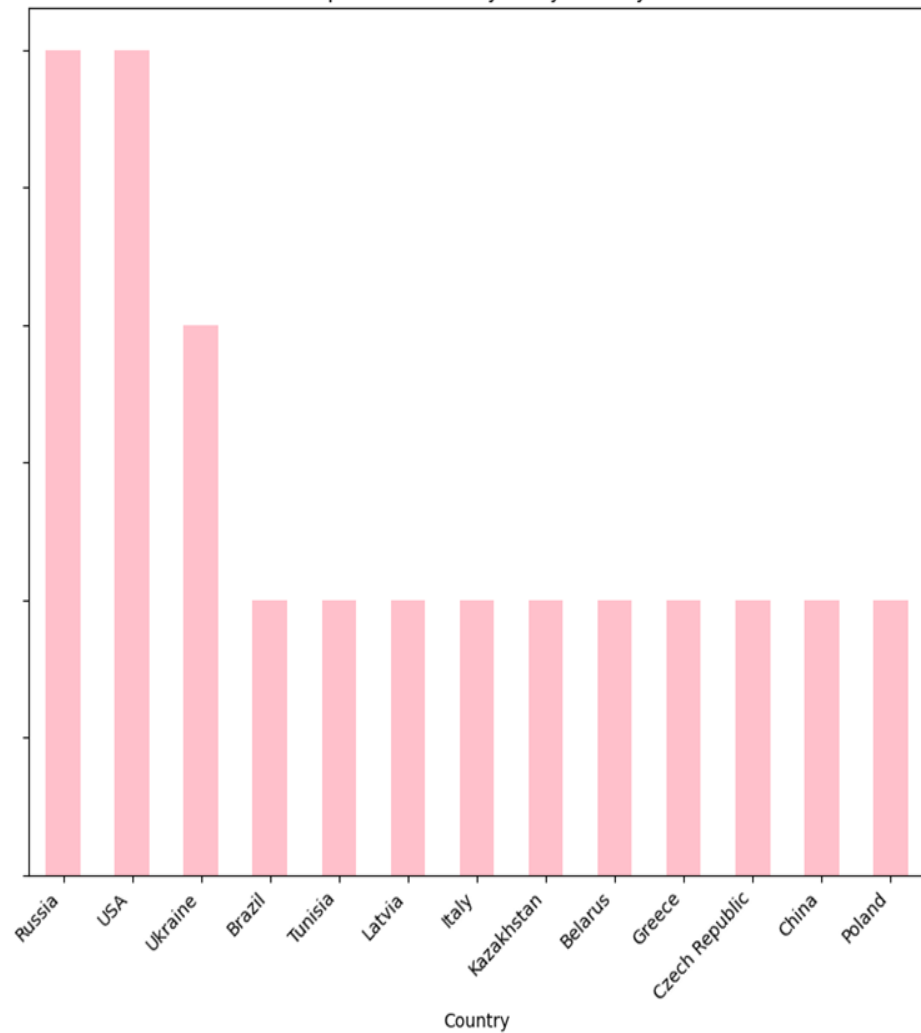
Countries with the maximum number of top 20 male players: USA, Russia

Countries with the maximum number of top 20 female players: Russia, USA

Top 20 Male Players by Country



Top 20 Female Players by Country



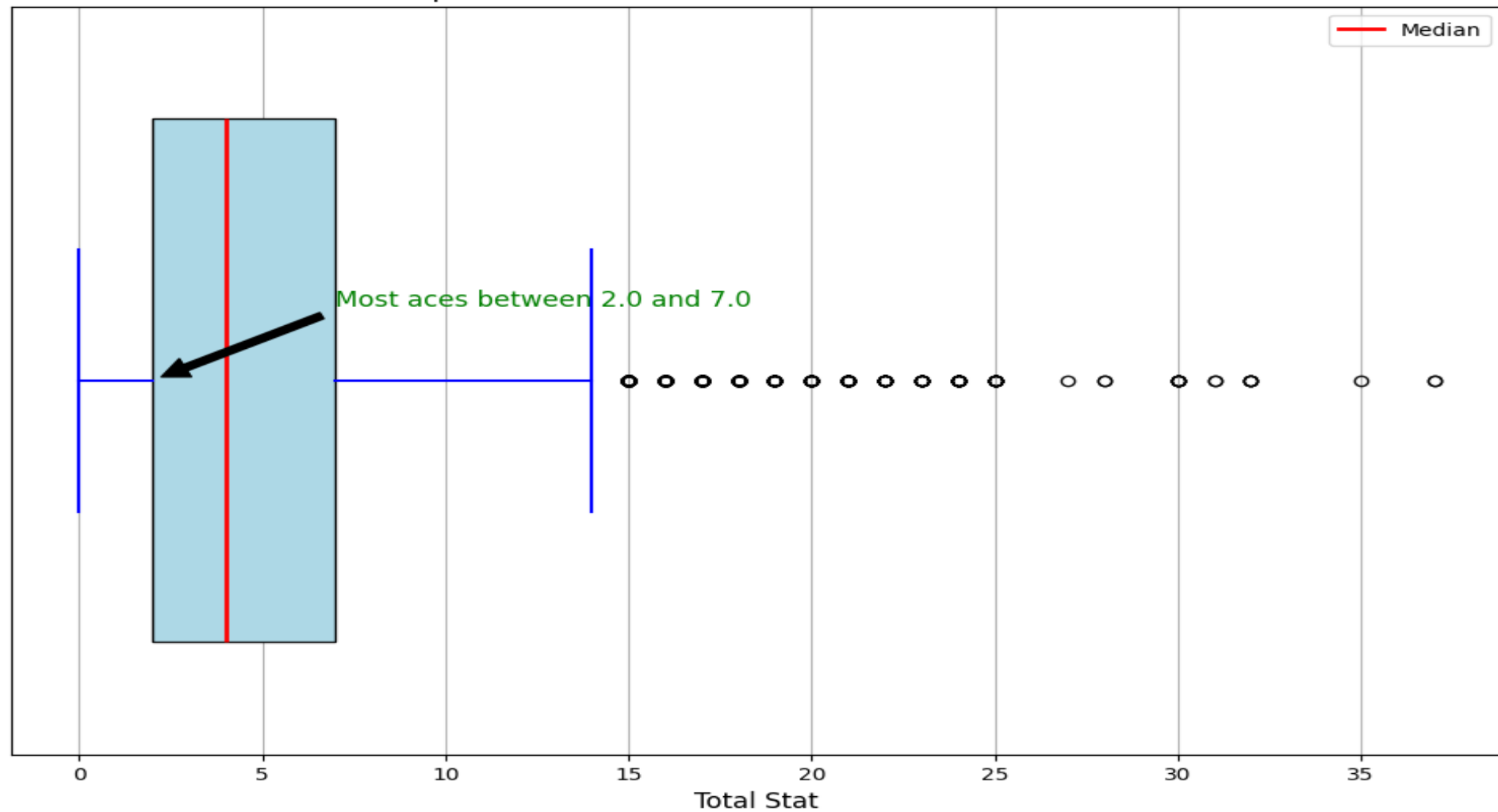
7. WHAT IS THE AVERAGE NUMBER OF ACES PER MATCH?

```
Mean of total_stat: 4.88
```

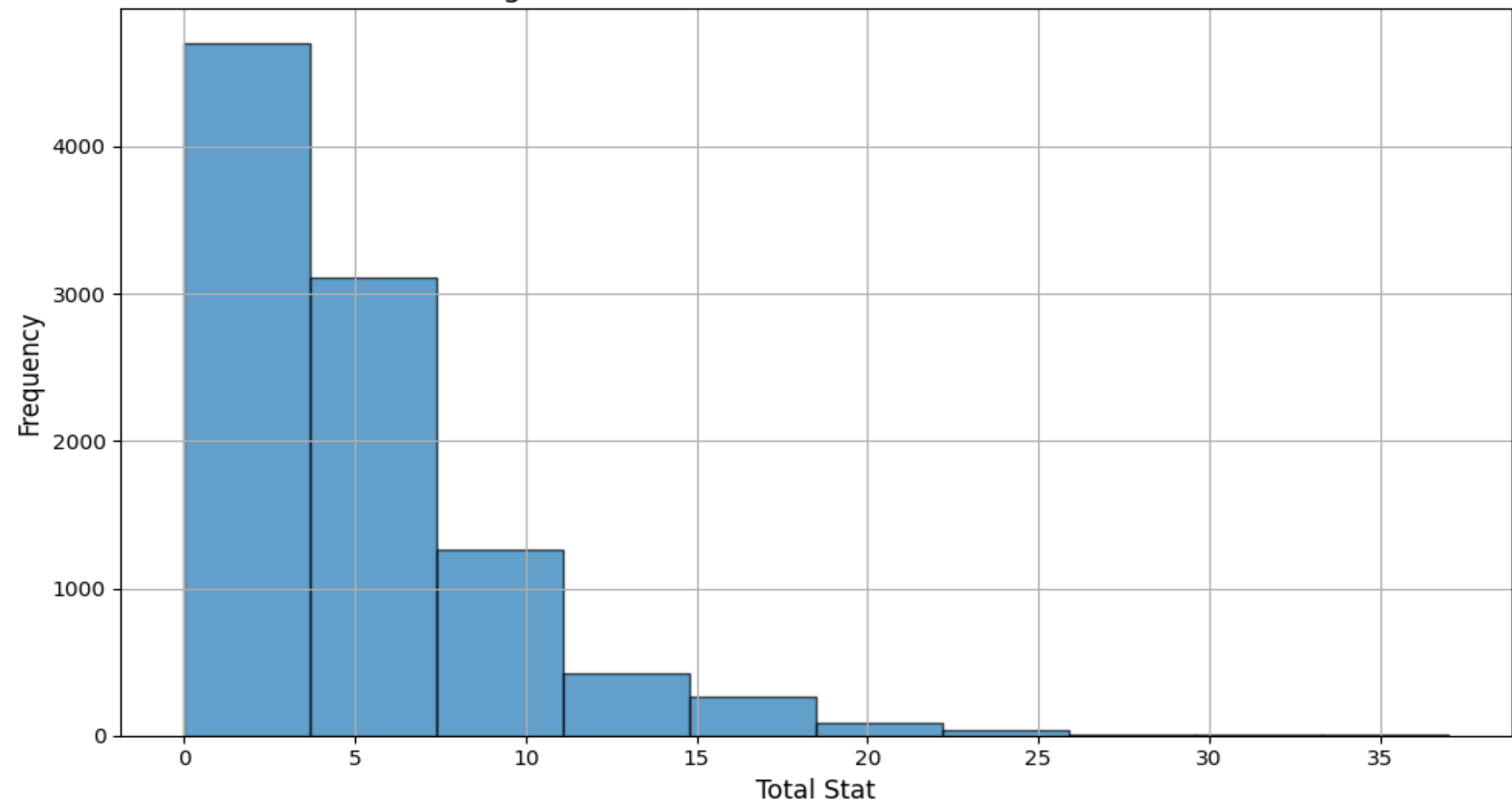
```
Median of total_stat: 4.00
```

```
the average number of aces per match is 4.88
```

Boxplot of Total Stat for "aces" and Period="ALL"



Histogram of Total Stat for "aces" and Period="ALL"

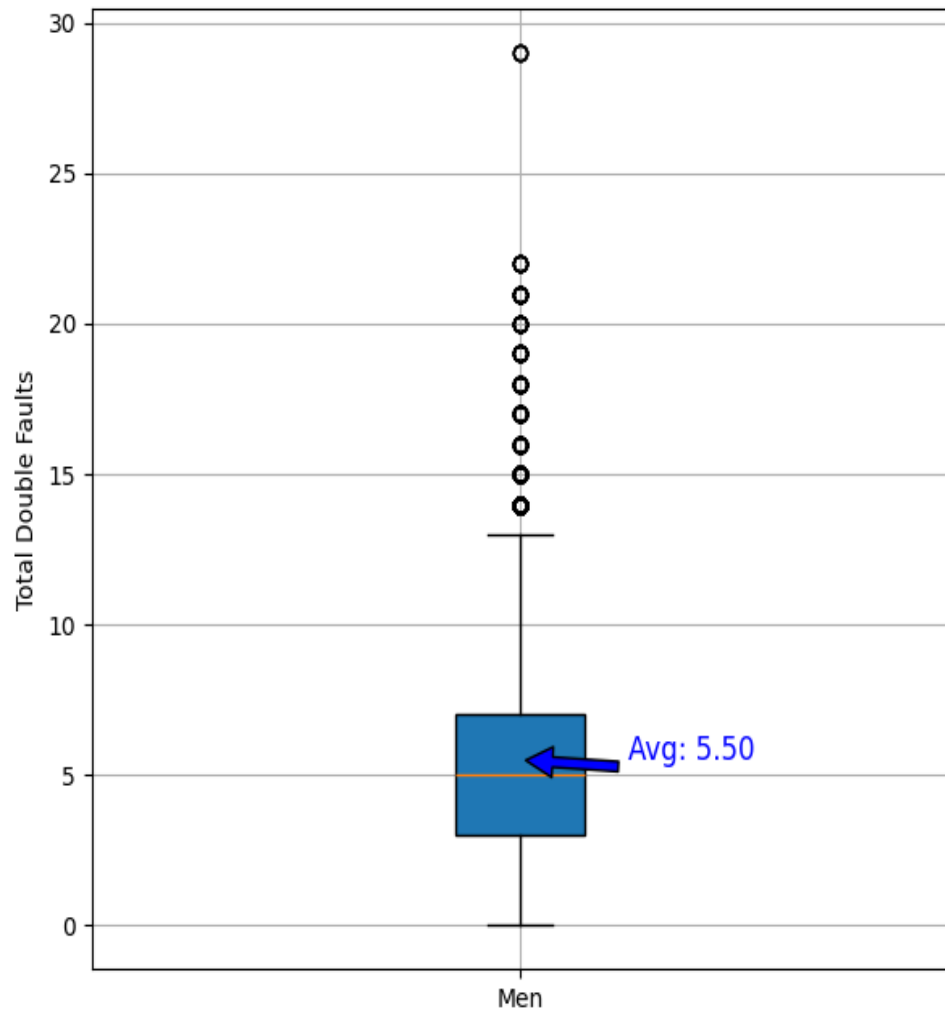


8. IS THERE A DIFFERENCE IN THE NUMBER OF DOUBLE
FAULTS BASED ON GENDER?

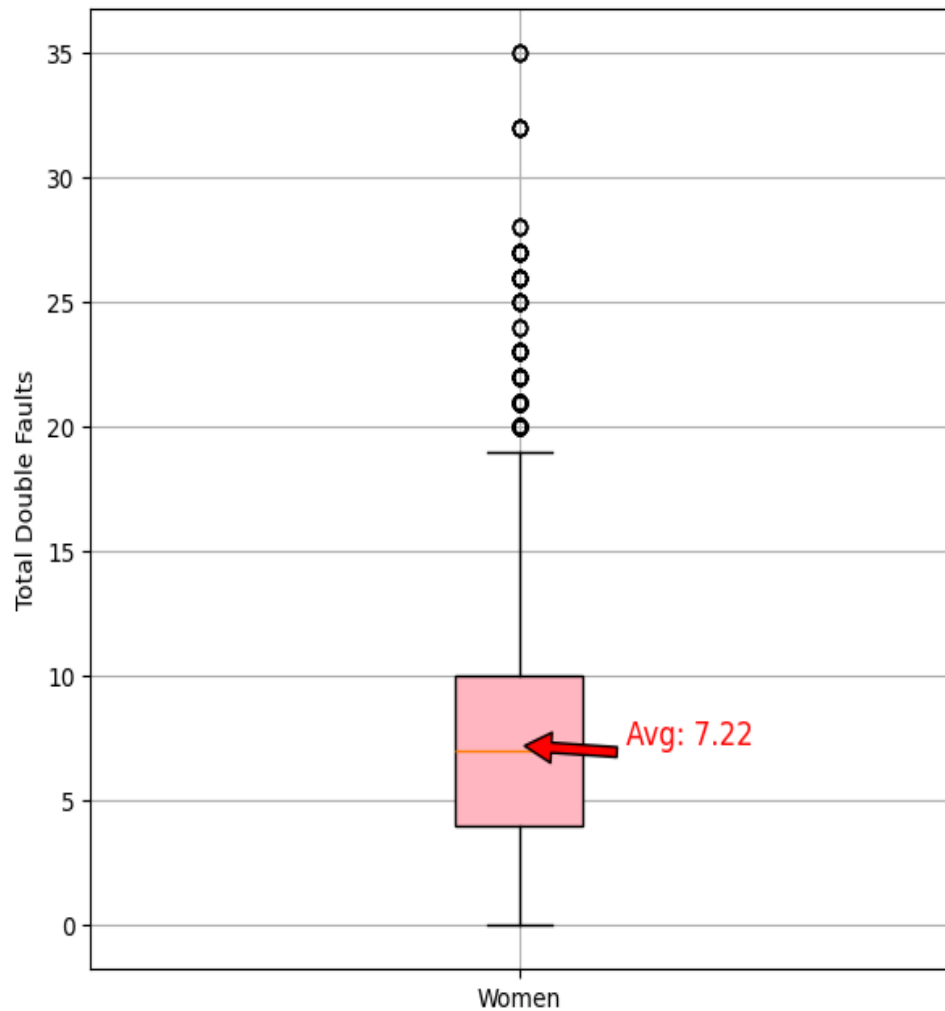
The average double faults of men is 5.503915246430217

The average double faults of women is 7.221408589764686

Double Faults - Men



Double Faults - Women



9. WHICH PLAYER HAS WON THE MOST TOURNAMENTS IN
A SINGLE MONTH?



The player with the most wins is Uchijima M. with 3 wins.

10. IS THERE A CORRELATION BETWEEN A PLAYER'S HEIGHT
AND THEIR RANKING?

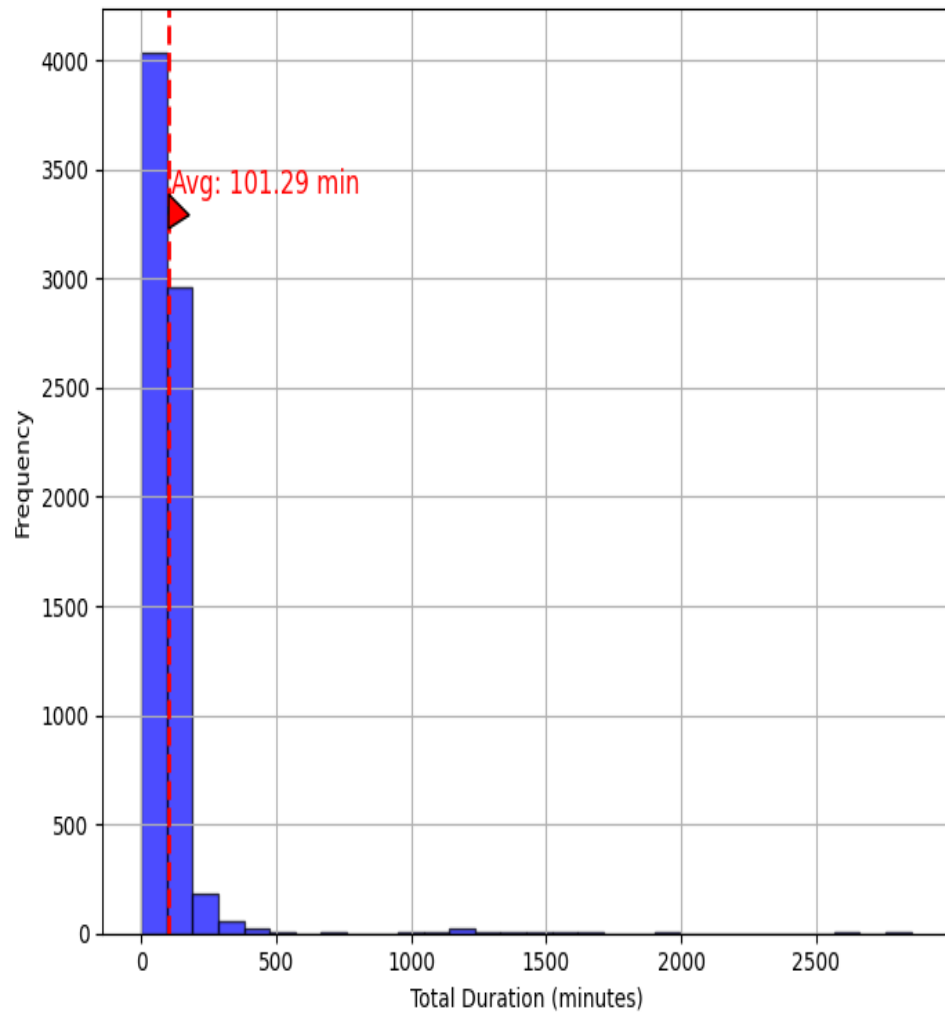
The correlation between height and ranking is: 0.08609144624811947

This correlation is categorized as: Very weak or no correlation

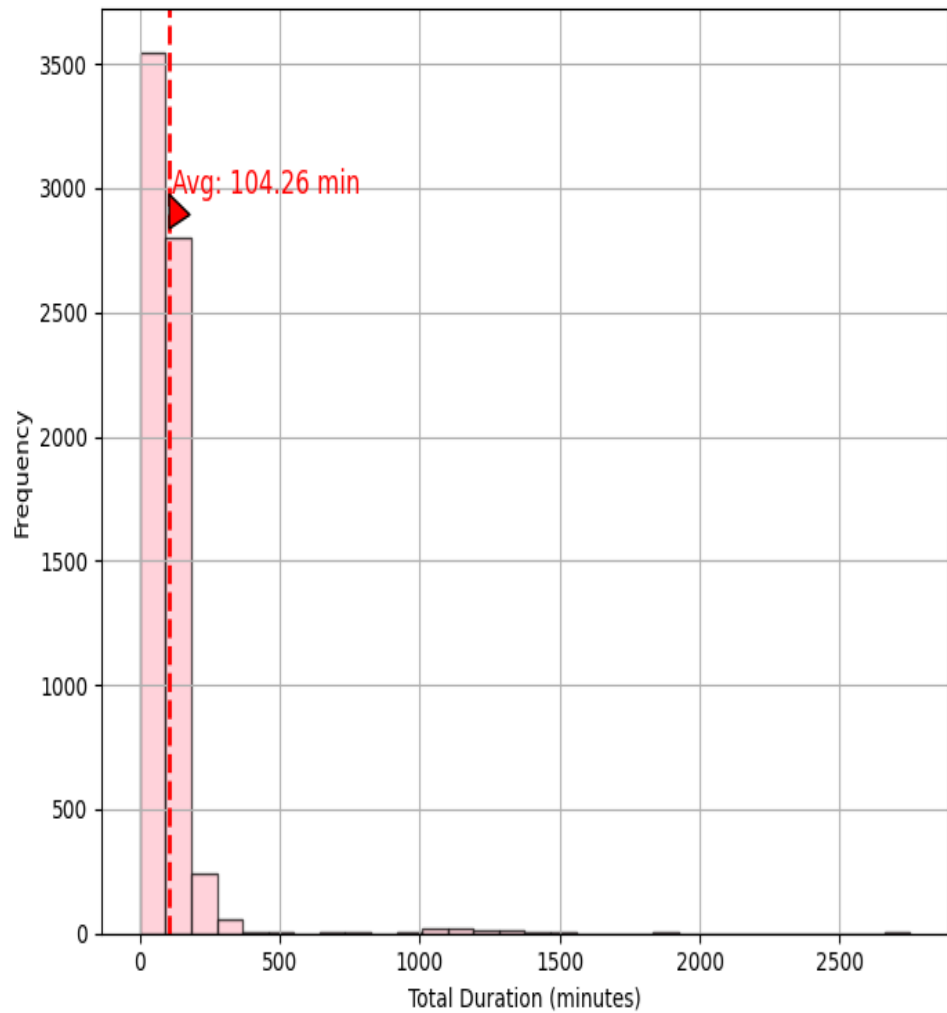
11. WHAT IS THE AVERAGE DURATION OF MATCHES?

The average duration of matches for men is 101.29 minutes.
The average duration of matches for women is 104.26 minutes.

Match Duration Distribution - Men



Match Duration Distribution - Women



12-WHAT IS THE AVERAGE NUMBER OF GAMES PER SET IN MEN'S MATCHES COMPARED TO WOMEN'S MATCHES?

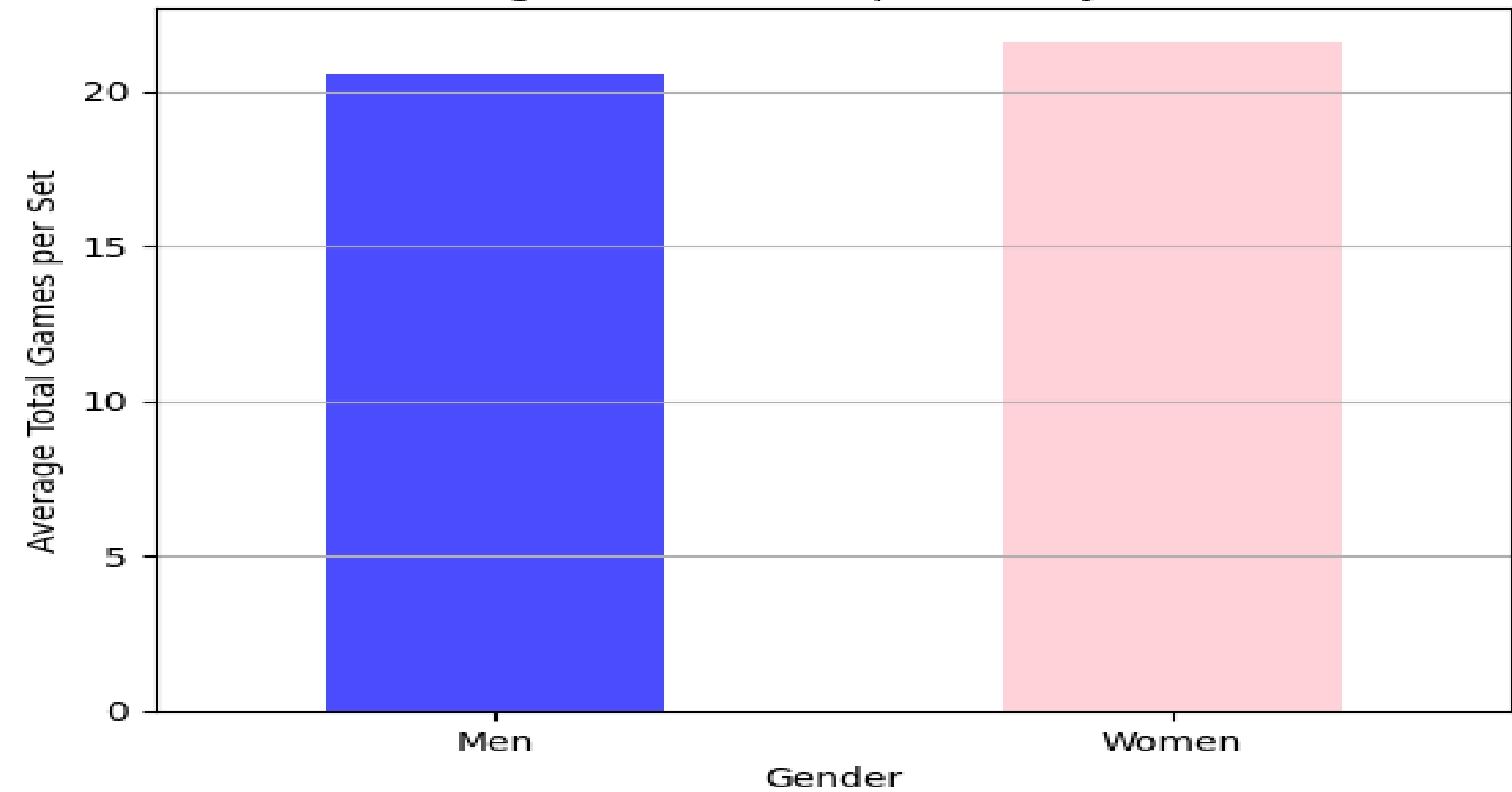
Average total games per set by gender:

gender

F 20.512120

M 21.599032

Average Total Games per Set by Gender



Unique values in 'plays' column:

```
['left-handed' 'right-handed' nan 'ambidextrous']
```

Number of missing or NaN values in 'plays' column: 1323

Initial Distribution of Left-handed vs Right-handed players:

plays

unknown	1323
---------	------

right-handed	908
--------------	-----

left-handed	120
-------------	-----

ambidextrous	1
--------------	---

Name: count, dtype: int64

Known values distribution for imputation:

plays

right-handed	0.882410
--------------	----------

left-handed	0.116618
-------------	----------

ambidextrous	0.000972
--------------	----------

Name: proportion, dtype: float64

Distribution of Left-handed vs Right-handed players after imputation:

plays

right-handed	2061
--------------	------

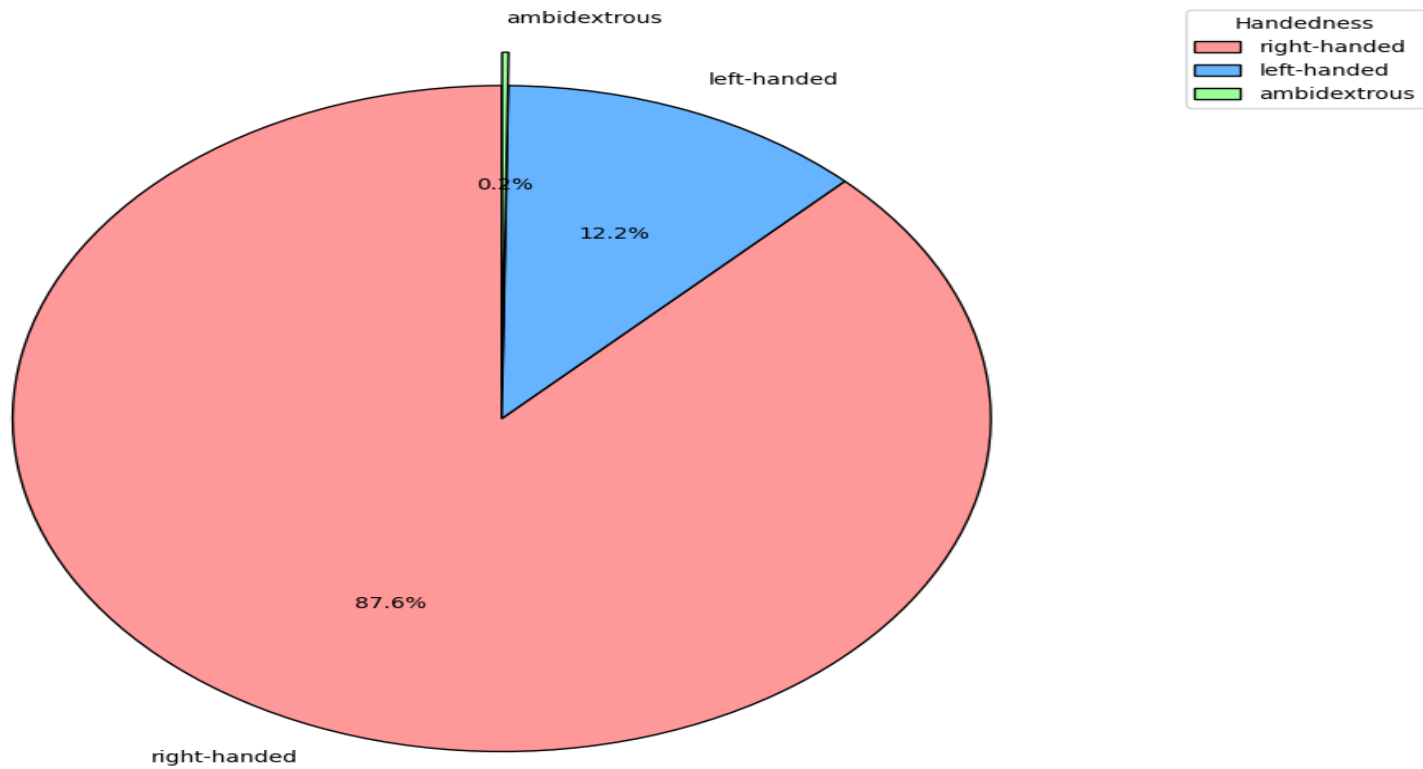
left-handed	286
-------------	-----

ambidextrous	5
--------------	---

Name: count, dtype: int64

13. WHAT IS THE DISTRIBUTION OF LEFT- HANDED VERSUS RIGHT- HANDED PLAYERS?

Distribution of Left-handed vs Right-handed Players



Ground Type Counts:

ground_type

Red clay 14545

Hardcourt outdoor 4712

Hardcourt indoor 257

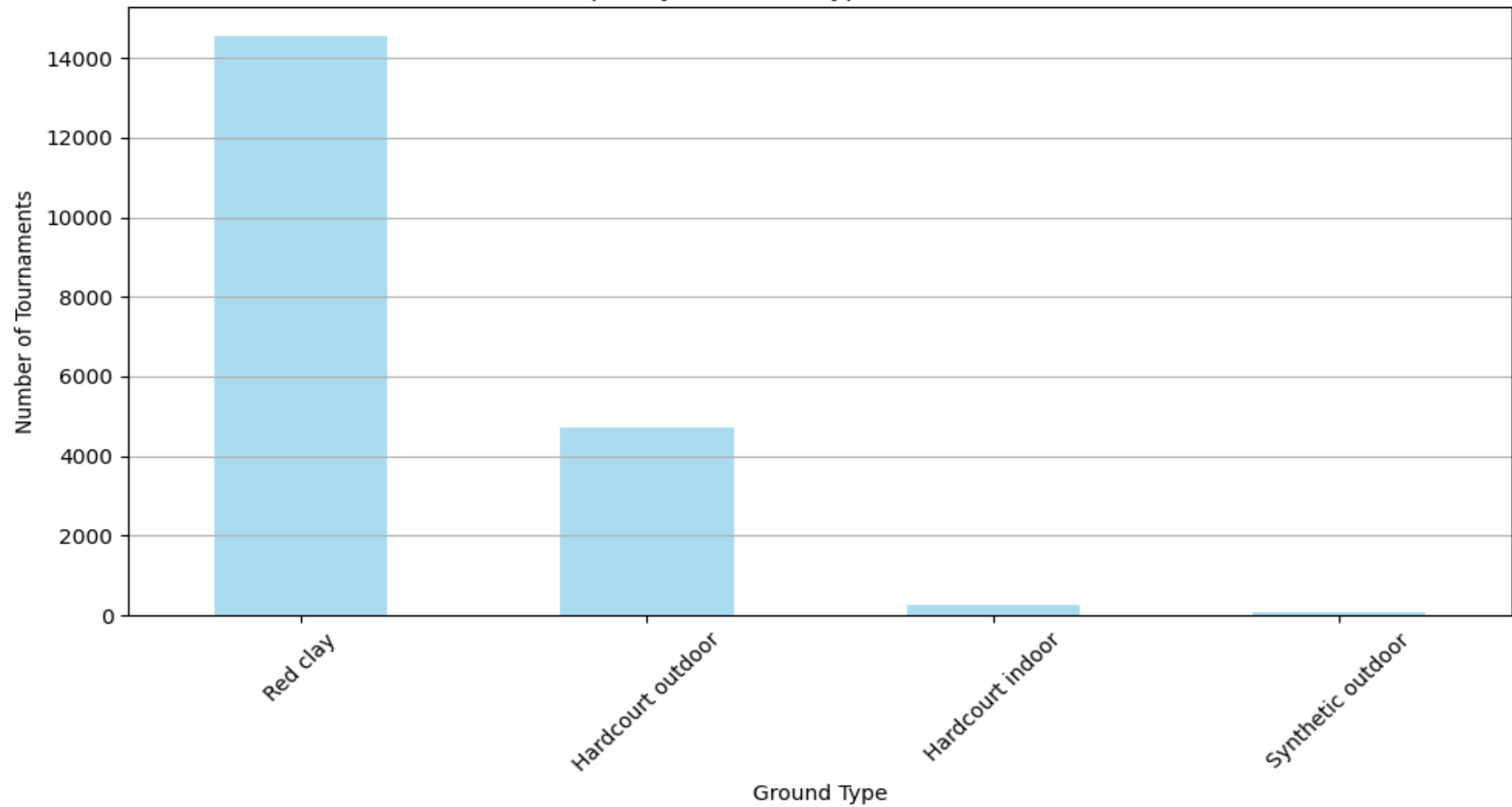
Synthetic outdoor 94

Name: count, dtype: int64

14. WHAT IS THE MOST
COMMON TYPE OF SURFACE
USED IN TOURNAMENTS?



Frequency of Ground Types in Tournaments



15. HOW MANY DISTINCT COUNTRIES ARE REPRESENTED
IN THE DATASET?

Number of distinct countries represented in the dataset: 97

16. WHICH PLAYER HAS THE HIGHEST WINNING PERCENTAGE AGAINST TOP 10 RANKED OPPONENTS?

Player: Griekspoor T. | Winning Percentage: 100.00% | Matches Played: 3

Player: Osaka N. | Winning Percentage: 100.00% | Matches Played: 3

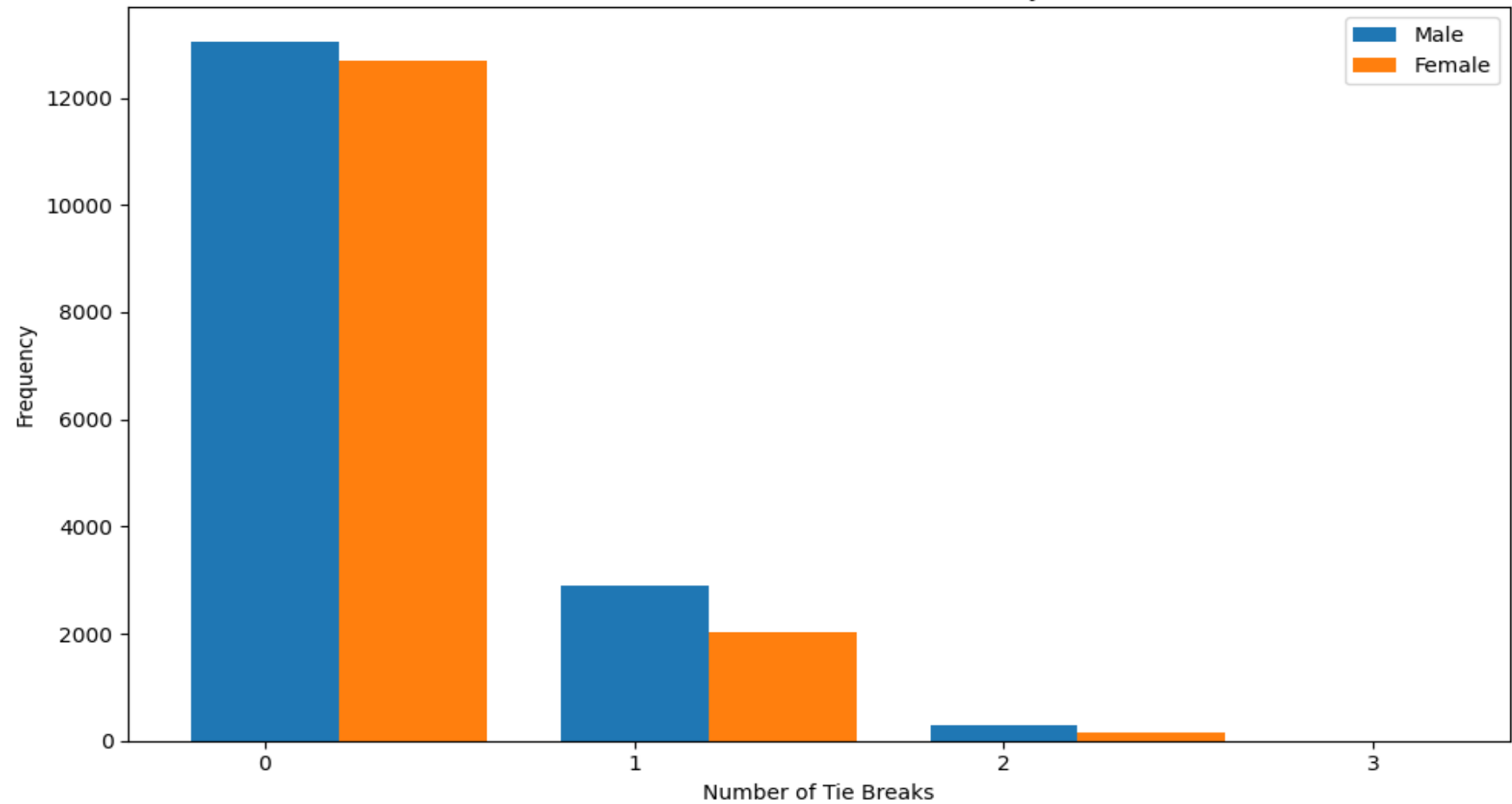
Player: Ruud C. | Winning Percentage: 100.00% | Matches Played: 3

17. WHAT IS THE AVERAGE NUMBER OF BREAKS OF SERVE PER MATCH?

Typically, the median number of tie breaks in a tennis match for males is 0.0.

Typically, the median number of tie breaks in a tennis match for females is 0.0.

Number of Tie Breaks in Tennis Matches by Gender

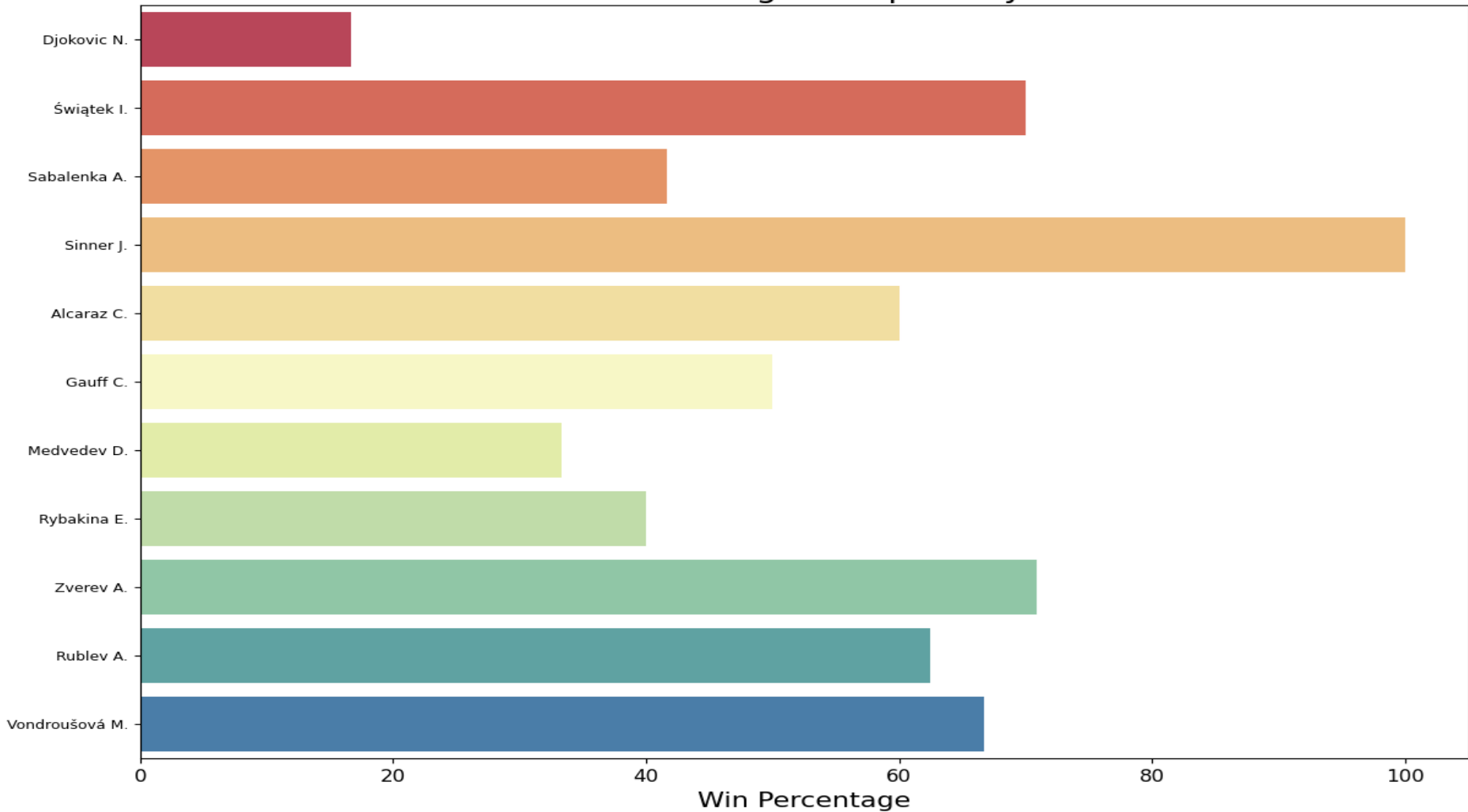


	name	current_rank	wins	games_played	win_percentage
0	Djokovic N.	1.0	1	6	16.666667
1	Świątek I.	1.0	7	10	70.000000
2	Sabalenka A.	2.0	5	12	41.666667
3	Sinner J.	2.0	4	4	100.000000
4	Alcaraz C.	3.0	3	5	60.000000
5	Gauff C.	3.0	4	8	50.000000
6	Medvedev D.	4.0	2	3	66.666667
7	Rybakina E.	4.0	2	5	40.000000
8	Zverev A.	4.0	6	8	75.000000
9	Medvedev D.	5.0	0	3	0.000000
10	Zverev A.	5.0	2	3	66.666667
11	Rublev A.	6.0	5	8	62.500000
12	Vondroušová M.	6.0	4	6	66.666667

18. WHAT IS PERCENTAGE
OF WINS AMONG THE
TOP 6 RANKED PLAYERS?

Win Percentage of Top 6 Players

Player Name



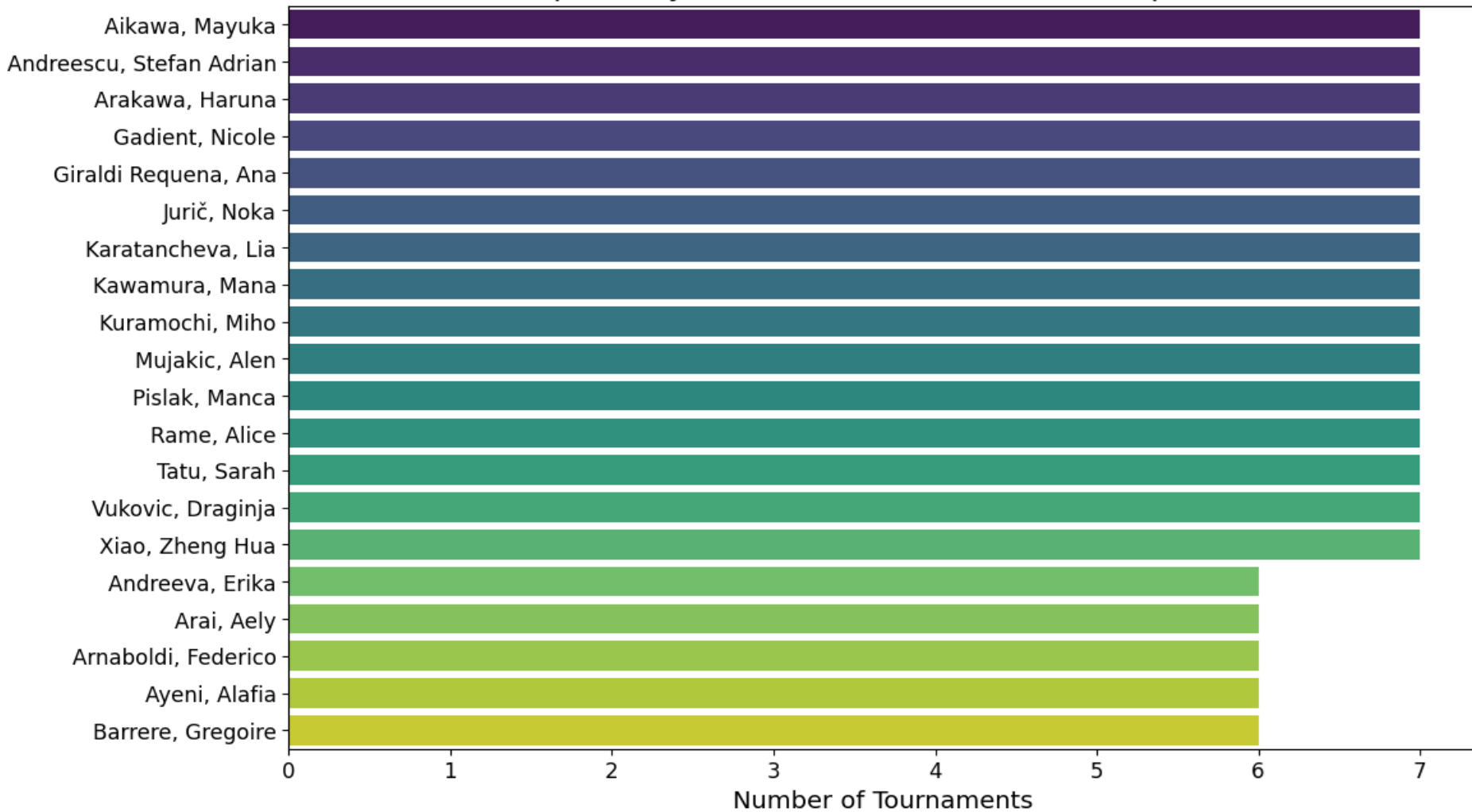
19. WHO WINS THE FINAL GAME IN EACH TOURNAMENTS?

	match_id	name_x	Winner	h/a	gender	round_id	name_y	tournament_name
0	12303219	Snigur D.	0	0	F	29	Final	Lopota, Singles Main, W-ITF-GEO-02A
1	12261565	Świątek I.	0	0	F	29	Final	Madrid, Spain
2	12276771	Schoolkate T.	0	0	M	29	Final	Guangzhou, China
3	12305661	Agwi M.	0	0	M	29	Final	Antalya, Singles Main, M-ITF-TUR-17A
4	12305681	El Aouni A.	0	0	F	29	Final	Antalya, Singles Main, W-ITF-TUR-17A
5	12305710	Searle H.	0	0	M	29	Final	Nottingham, Singles Main, M-ITF-GBR-10A
6	12305721	Wiskandt M.	0	0	M	29	Final	Monastir, Singles Main, M-ITF-TUN-30A
7	12305734	Yamazaki I.	0	0	F	29	Final	Hammamet, Singles Main, W-ITF-TUN-22A
8	12305784	Back D.	0	0	F	29	Final	Monastir, Singles Main, W-ITF-TUN-27A
9	12305918	Kartal S.	0	0	F	29	Final	Nottingham, Singles Main, W-ITF-GBR-13A
10	12305966	Clarke J.	0	0	M	29	Final	S. Margherita di Pula, Singles Main, M-ITF-ITA...
11	12305991	Bielinskyi V.	0	0	M	29	Final	Osijek, Singles Main, M-ITF-CRO-06A
12	12306132	Herea C.	0	0	F	29	Final	Kursumlijska Banja, Singles Main, W-ITF-SRB-03A
13	12306333	Kostic T.	0	0	F	29	Final	Yecla, Singles Main, W-ITF-ESP-20A
14	12306352	Gueymard Wayenburg S.	0	0	M	29	Final	Kursumlijska Banja, Singles Main, M-ITF-SRB-03A
15	12261532	Auger-Aliassime F.	0	0	M	29	Final	Madrid, Spain
16	12276840	Tabilo A.	0	0	M	29	Final	Aix en Provence, France
17	12276867	Navone M.	0	0	M	29	Final	Cagliari, Italy
18	12276932	Paquet C.	0	0	F	29	Final	Saint-Malo, France
19	12306970	Joint M.	0	0	F	29	Final	Bonita Springs, Singles Main, W-ITF-USA-13A

20. TOP 20 PLAYERS WITH MOST TOURNAMENT PARTICIPATION

Top 20 Players with Most Tournament Participation

Player Name



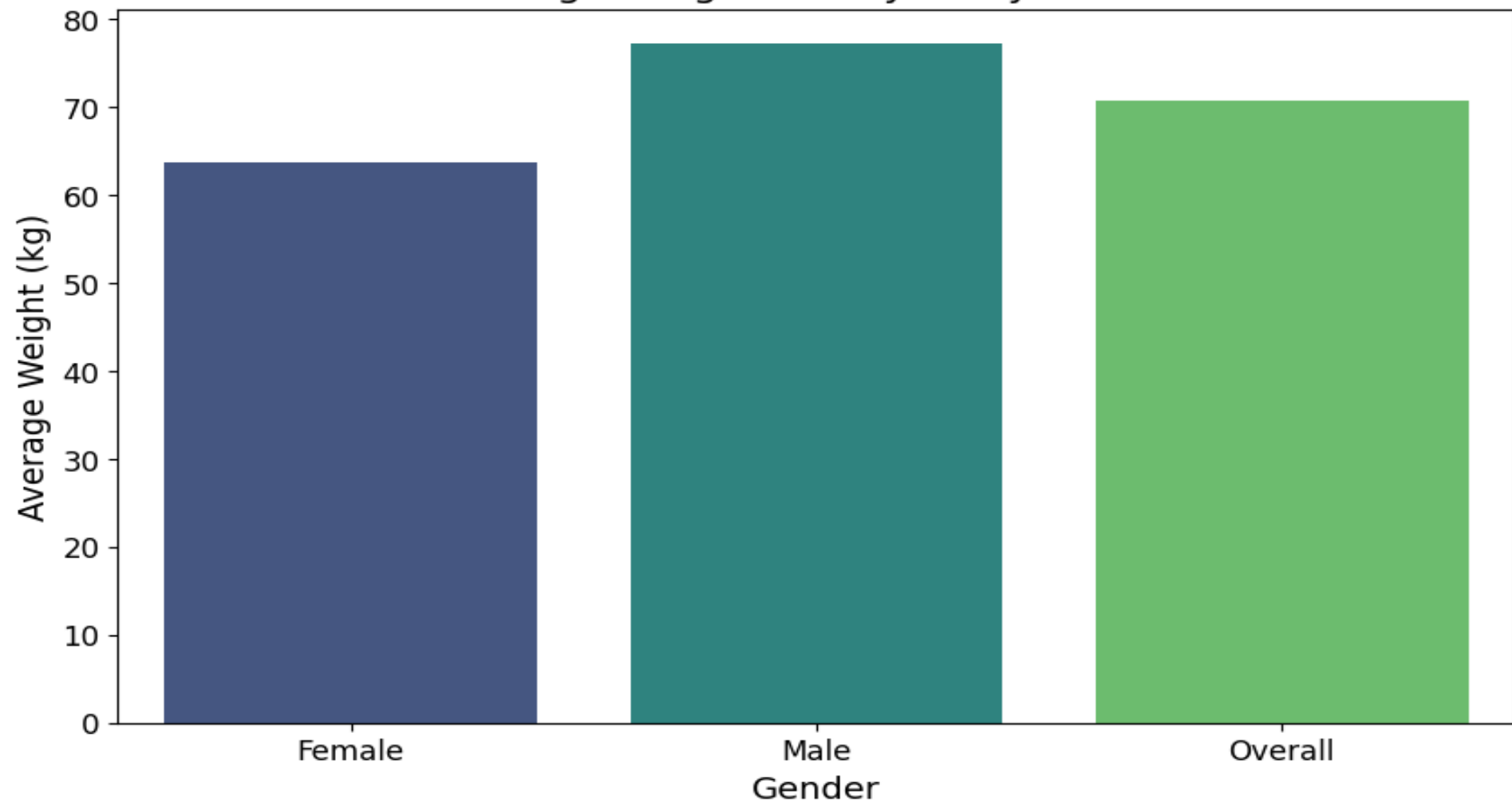
21. WHAT IS THE AVERAGE WEIGHT OF THE PLAYERS?

The average weight of female players is 63.74489795918368

The average weight of male players is 77.2742268041237

The average weight of players is 70.72573188812376

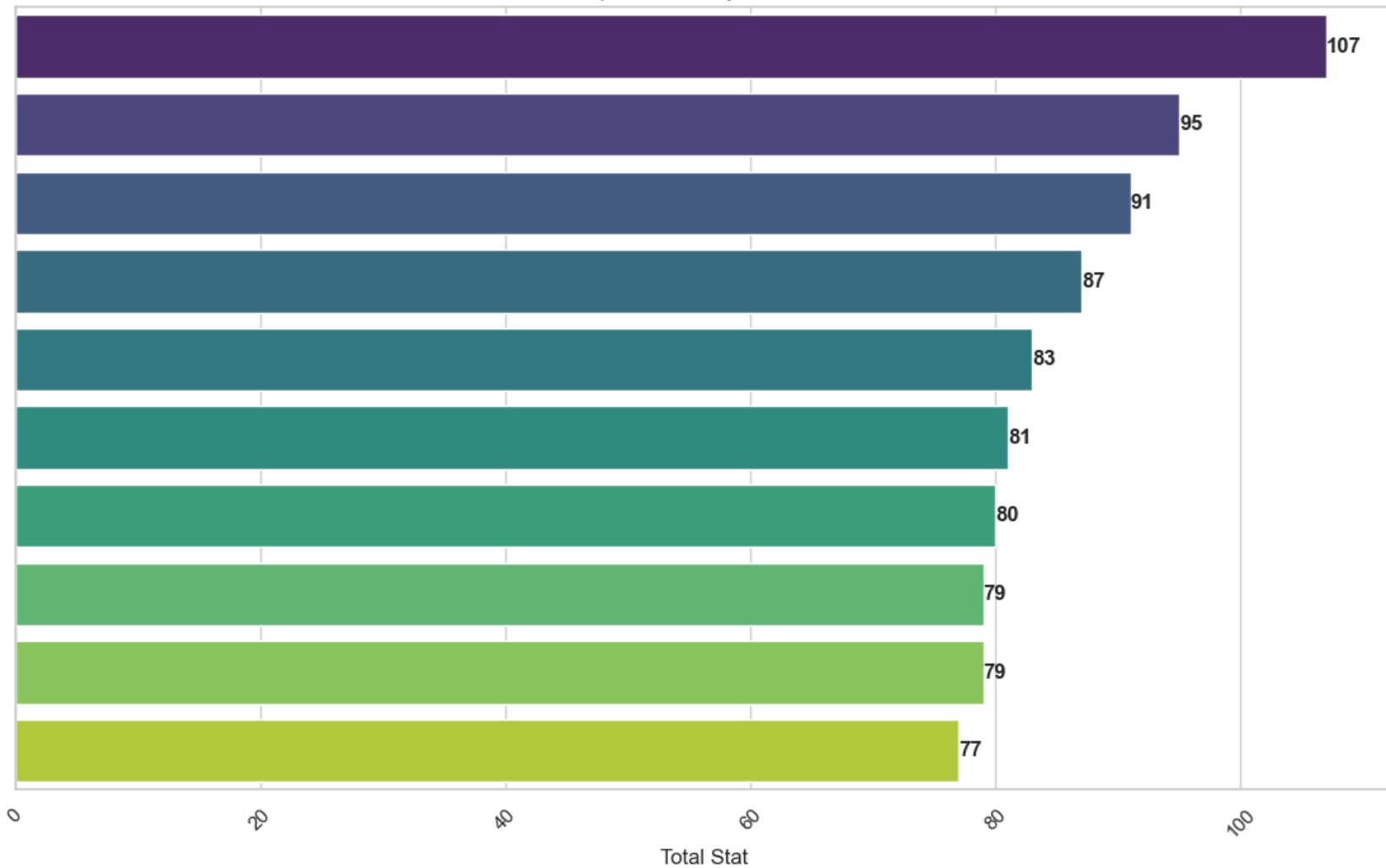
Average Weight of Players by Gender



22.FIND TOP 10 PLAYERS WITH MOST ACES POINTS IN
THIS MONTH

Top 10 Teams by Total Stat

Team Name



23. WHAT IS THE CORRELATION BETWEEN THOSE WHOSE RANK IS AMONG THE TOP 10 AND THE NUMBER OF THEIR ACES?

```
Correlation for males: 0.2357642931489974
Correlation strength for total data: weak
Correlation for females: -0.45274081020048734
Correlation strength for total data: moderate
Total correlation: 0.05613886420404405
Correlation strength for total data: weak
```

24. WHICH COUNTRY HOSTED THE MOST GAMES AND HOW
MANY GAMES?

Country with the maximum games hosted: France

Number of games hosted: 803

25. WHAT HAS BEEN THE WIN PERCENTAGE OF PLAYERS WHO PLAYED IN THEIR OWN COUNTRY AND HAD MORE THAN 10 MATCHES PLAYED?

	index	name	gender	num_wins	num_games	win_percentage
0	119	Chirita I.	M	9	12	75.000000
1	551	Poljičak M.	M	8	13	61.538462
2	672	Sun F.	M	9	16	56.250000
3	147	Damas M.	M	6	11	54.545455
4	34	Bai Y.	M	7	13	53.846154
5	60	Boitan G.	M	7	14	50.000000
6	462	Mo Y.	M	6	12	50.000000
7	664	Stankovic A.	F	7	14	50.000000
8	709	Urhobo A.	F	7	15	46.666667
9	141	Cui J.	M	4	11	36.363636
10	521	Passaro F.	M	4	11	36.363636
11	639	Senic N.	F	4	11	36.363636

THANKS FOR YOUR TIME