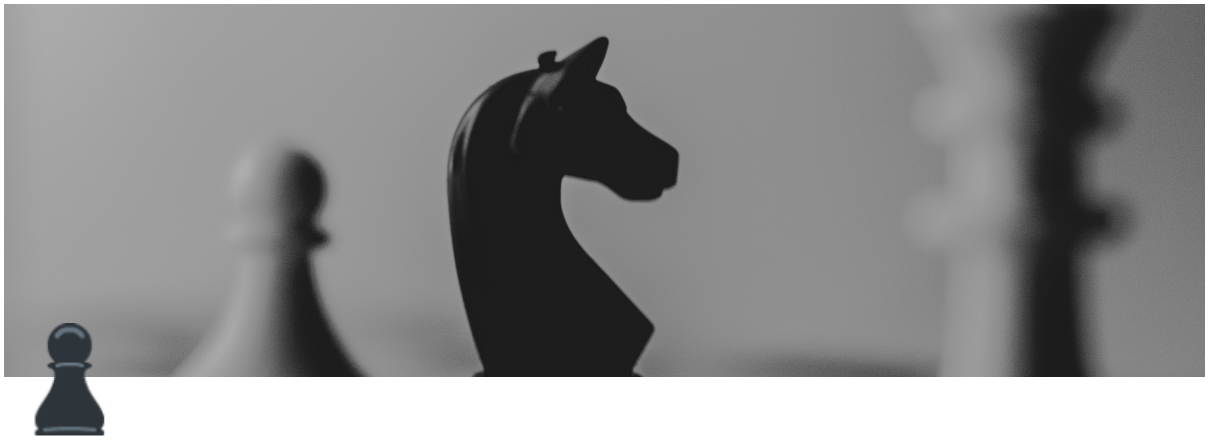


Analysis of Top Chess Players

> Chaitanya Raj Dixit 18/94067

> Jagriti Verma 18/94058

Description of Dataset

Overview

Relative chess skill level is measured by Elo rating. The International Chess Federation, FIDE, acts as the governing body for international chess. FIDE assigns Elo ratings for three speeds of chess (in ascending order): Standard, Rapid, and Blitz. They also award various titles, based on both rating and performance.

This dataset has data of the chess rankings in September 2020.

https://www.kaggle.com/hunterklotzburwell/top-chess-players

Selected Subset

For our analysis, we have selected the topchesslist.csv table.

This dataset contains information on 746 active chess players whose Standard FIDE ratings were at least 2500 (as of September 2020).

The 2500 minimum was chosen to match the peak rating requirement for the Grandmaster (GM) title.

The dataset has the following columns :

- Name: Name of the Player
- Title: Whether the player is a GM or an IM
- Country: The player's country of origin
- Standard Rating: FIDE Rating of Standard Format (thinking time >60) of Chess
- Rapid Rating: FIDE Rating of Rapid Format (thinking time >10 and <60) of Chess
- Blitz Rating: FIDE Rating of Blitz Format (thinking time <10) of Chess
- Birth Year: Birth year of the player

Objective

By aggregating the data according to the age and country of the players, we can see how much a country dominates the sport and which age groups occupy the top positions.

We would also visualize our results to make them more intuitive. For this, we'll use pie charts, bar graphs, and other plots included in the ggplot2 and base package.
The following section contains our analysis, observations, and visualizations.

Descriptive Analysis

1. Dataset

```
players <- read.csv("topchesslist.csv")
print("DATASET")
print(players)
```

```
[1] "Top Chess Players DATASET"
                            Name  Title Country Standard Rapid Blitz Birth.Year
1                  Carlsen, Magnus     GM     NOR     2863  2881  2886       1990
2                 Caruana, Fabiano     GM     USA     2835  2773  2711       1992
3                     Ding, Liren     GM     CHN     2791  2836  2788       1992
4             Nepomniachtchi, Ian     GM     RUS     2784  2778  2785       1990
5          Vachier-Lagrave, Maxime     GM     FRA     2778  2860  2822       1990
6              Grischuk, Alexander     GM     RUS     2777  2784  2765       1983
7                   Aronian, Levon     GM     ARM     2773  2778  2739       1982
8                       So, Wesley     GM     USA     2770  2741  2816       1993
9               Radjabov, Teimour     GM     AZE     2765  2758  2757       1987
10           Mamedyarov, Shakhriyar     GM     AZE     2764  2761  2716       1985
11                    Giri, Anish     GM     NED     2764  2731  2752       1994
12                     Wang, Hao     GM     CHN     2763  2750  2712       1989
13               Rapport, Richard     GM     HUN     2760  2727  2759       1996
14         Dominguez Perez, Leinier     GM     USA     2758  2786  2654       1983
15              Anand, Viswanathan     GM     IND     2753  2751  2785       1969
16             Duda, Jan-Krzysztof     GM     POL     2753  2774  2799       1998
17                Karjakin, Sergey     GM     RUS     2752  2709  2765       1990
18                Nakamura, Hikaru     GM     USA     2736  2829  2900       1987
19                Topalov, Veselin     GM     BUL     2735  2707  2667       1975
20                        Wei, Yi     GM     CHN     2732  2752  2686       1999
21            Harikrishna, Pentala     GM     IND     2732  2705  2614       1986
22               Firouzja, Alireza     GM     FID     2728  2703  2770       2003
23          Vidit, Santosh Gujrathi     GM     IND     2726  2636  2735       1994
24               Andreikin, Dmitry     GM     RUS     2726  2740  2782       1990
25                  Svidler, Peter     GM     RUS     2723  2742  2754       1976
26                Vitiugov, Nikita     GM     RUS     2722  2616  2673       1987
27              Wojtaszek, Radoslaw     GM     POL     2717  2703  2730       1987
28              Artemiev, Vladislav     GM     RUS     2716  2769  2783       1998
29          Vallejo Pons, Francisco     GM     ESP     2710  2656  2615       1982
30                      Yu, Yangyi     GM     CHN     2709  2738  2808       1994
31                   Xiong, Jeffery     GM     USA     2709  2730  2724       2000
32                   Le, Quang Liem     GM     VIE     2709  2744  2690       1991
```

```
print("Type of dataset")
print(class(players))
print(mode(players))
print(typeof(players))

obs <- nrow(players)
var <- ncol(players)
print("no of observations")
print(obs)
print("no of variables")
print(var)
attach(players)
```

```
[1] "Type of Dataset : "
[1] "data.frame"
[1] "list"
[1] "list"
[1] "Number of observations : "
[1] 746
[1] "Number of variables : "
[1] 7
```

2. Analysis based on Standard Rating

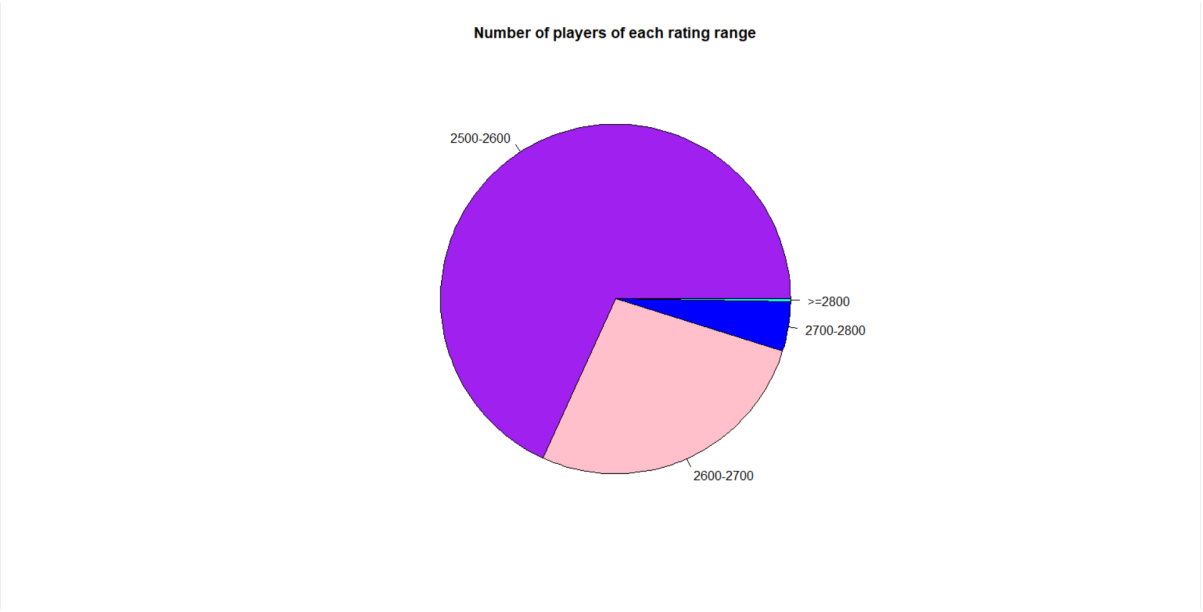
```
print("Info about Standard Rating")
print(summary(players$Standard))
```

```
[1] "Info about Standard Rating"
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2500    2525    2562    2577    2614    2863
```

```
print("Different range of Standard Ratings")
n <- c()
print("No. of players of rating 2500-2600")
n[1] <- length(Standard[Standard < 2600])
print(n[1])
print("No. of players of rating 2600-2700")
n[2] <- length(Standard[(Standard >= 2600) & (Standard < 2700)])
print(n[2])
print("No. of players of rating 2700-2800")
n[3] <- length(Standard[(Standard >= 2700) & (Standard < 2800)])
print(n[3])
print("No. of players of rating greater 2800")
n[4] <- length(Standard[Standard >= 2800])
print(n[4])
```

```
[1] "Different range of Standard Ratings"
[1] "No. of players of rating 2500-2600"
[1] 509
[1] "No. of players of rating 2600-2700"
[1] 201
[1] "No. of players of rating 2700-2800"
[1] 34
[1] "No. of players of rating greater 2800"
[1] 2
```

```
pie(n, c("2500-2600", "2600-2700", "2700-2800", ">=2800"), main = "Number of players of e
```



By observing this pie chart, we can conclude that only few (2) players have breached the 2800 Standard rating while most of the GM Ratings lie in the 2500-2600 bracket.

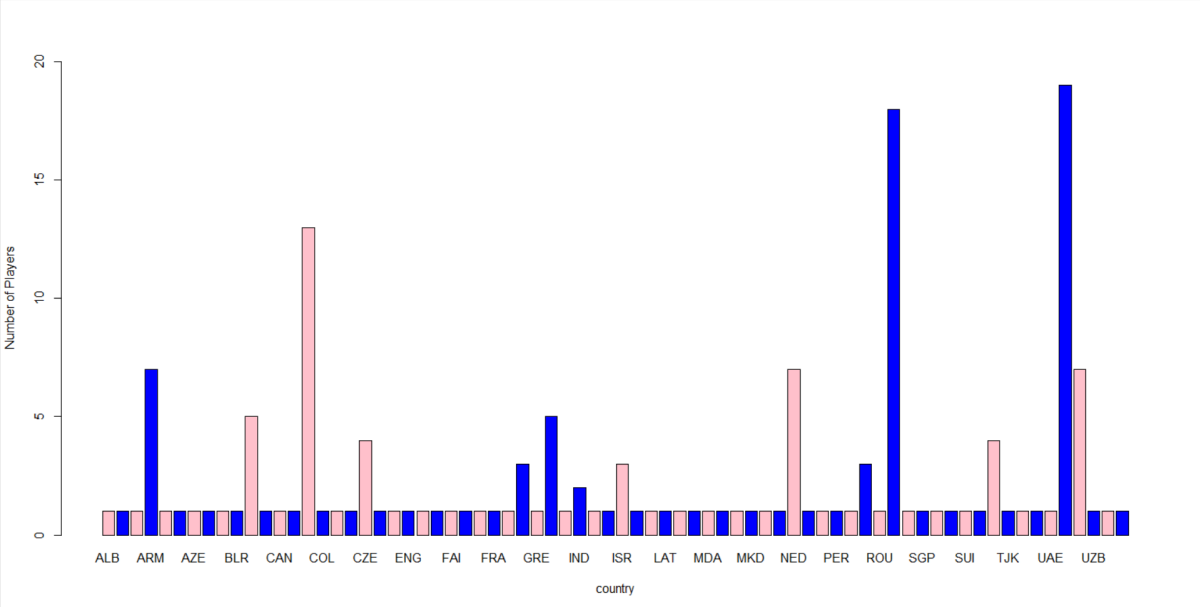
3. Analysis based on Country

```
# Number of players by Country
agg_country <- aggregate(Standard ~ (Country), players, function(x) which.max(table(x)))
print(agg_country)
```

	Country	Standard
1	ALB	1
2	ALG	1
3	ARG	1
4	ARM	7
5	AUS	1
6	AUT	1
7	AZE	1
8	BEL	1
9	BIH	1
10	BLR	1
11	BRA	5
12	BUL	1
13	CAN	1
14	CHI	1
15	CHN	13
16	COL	1
17	CRO	1
18	CUB	1
19	CZE	4
20	DEN	1
21	EGY	1
22	ENG	1
23	ESP	1
24	EST	1
25	FAI	1
26	FID	1
27	FIN	1
28	FRA	1
29	GEO	1
30	GER	3
31	GRE	1
32	HUN	5
33	INA	1
34	IND	2
35	IRI	1
36	ISL	1
37	ISR	3
38	ITA	1

39	KAZ	1
40	LAT	1
41	LTU	1
42	MAS	1
43	MDA	1
44	MEX	1
45	MGL	1
46	MKD	1
47	MNE	1
48	MYA	1
49	NED	7
50	NOR	1
51	PAR	1
52	PER	1
53	PHI	1
54	POL	3
55	ROU	1
56	RUS	18
57	SCO	1
58	SGP	1
59	SLO	1
60	SRB	1
61	SUI	1
62	SVK	1
63	SWE	4
64	TJK	1
65	TKM	1
66	TUR	1
67	UAE	1
68	UKR	19
69	USA	7
70	UZB	1
71	VEN	1
72	VIE	1

```
barplot(agg_country$Standard,
    width = 1, names.arg = agg_country$Country,
    horiz = FALSE, xlab = "country", ylab = "Number of Players", axes = TRUE,
    col = c("pink", "blue"), ylim = c(0, 20)
)
```

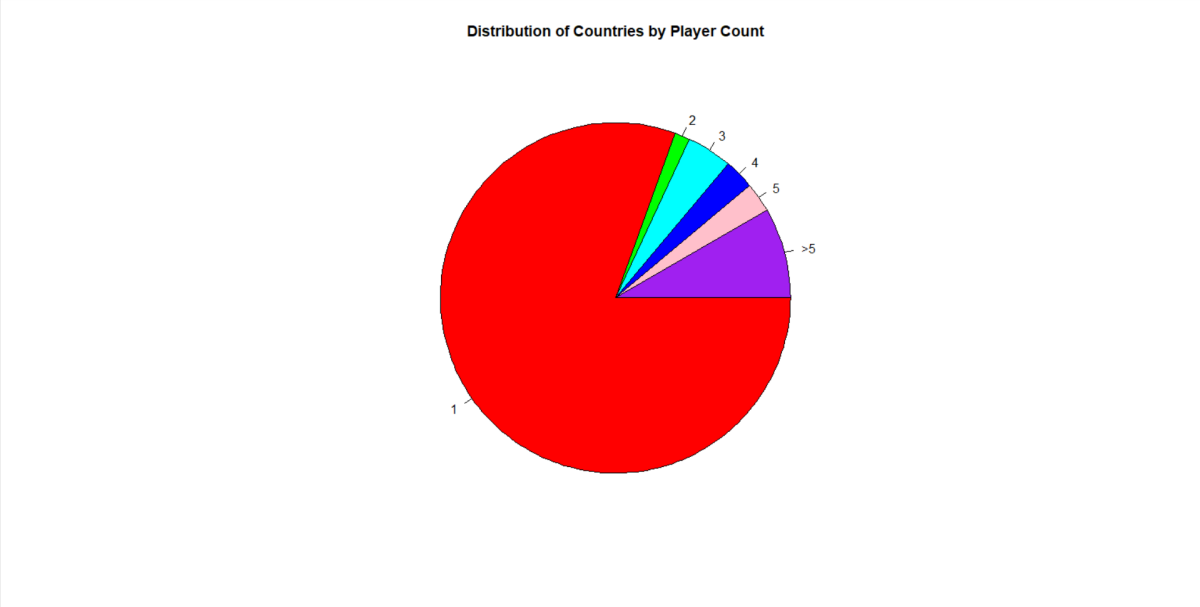



```
print("Different range of players by Country")
n <- c()
print("Countries with more than 5 players")
n[1] <- length(agg_country$Standard[agg_country$Standard > 5])
print(n[1])
print("Countries with 5 players")
n[2] <- length(agg_country$Standard[agg_country$Standard == 5])
print(n[2])
print("Countries with 4 players")
n[3] <- length(agg_country$Standard[agg_country$Standard == 4])
print(n[3])
print("Countries with 3 players")
n[4] <- length(agg_country$Standard[agg_country$Standard == 3])
print(n[4])
print("Countries with 2 players")
n[5] <- length(agg_country$Standard[agg_country$Standard == 2])
print(n[5])
print("Countries with only 1 player")
n[6] <- length(agg_country$Standard[agg_country$Standard == 1])
print(n[6])
```

By observing the bar graph, we can see that a select few countries (namely Ukraine(19), Russia(18), and China(13)) are dominating the top chess rankings.

```
[1] "Different range of players by Country"
[1] "Countries with more than 5 players"
[1] 6
[1] "Countries with 5 players"
[1] 2
[1] "Countries with 4 players"
[1] 2
[1] "Countries with 3 players"
[1] 3
[1] "Countries with 2 players"
[1] 1
[1] "Countries with only 1 player"
[1] 58
```

```
pie(n, c(">5", "5", "4", "3", "2", "1"), main = "Distribution of Countries by Player Cour
```



By observing this pie chart, we can see that most of the countries have only one player representing them with a rating above 2500 while only few countries have more than 5.

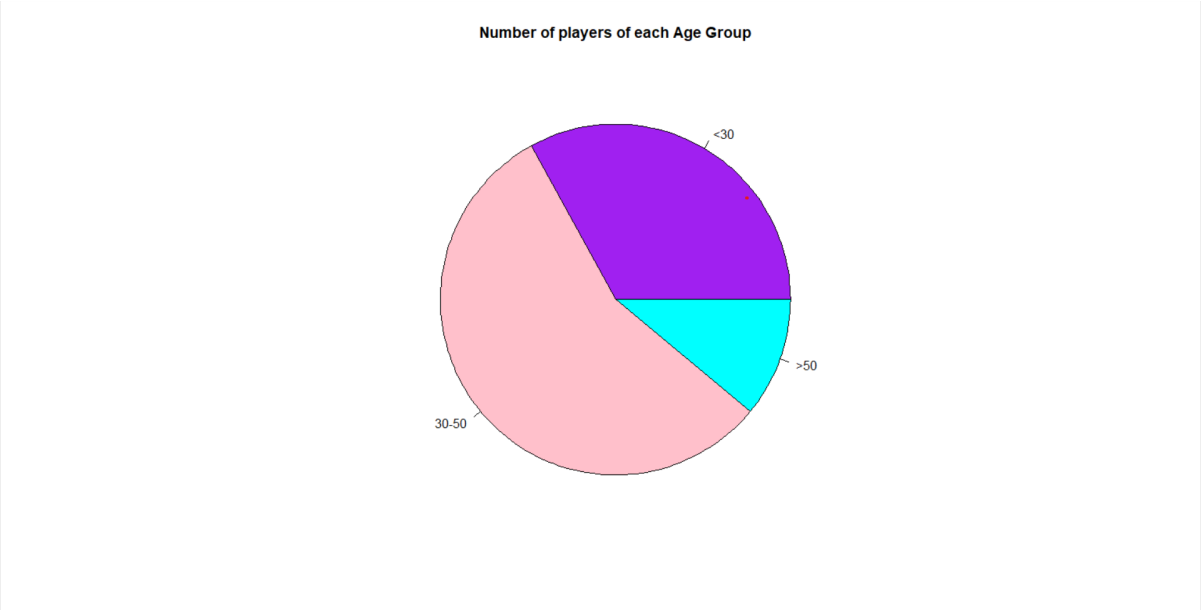
4. Analysis based on Age

```
players_age <- players
players_age$Birth.Year <- 2020 - players_age$Birth.Year
print("Different Age Range of Players")
n <- c()
print("Number of players of Age less than 30")
n[1] <- length(players_age$Birth.Year[players_age$Birth.Year < 30])
print(n[1])
```

```
print("Number of players of Age 30-50")
n[2] <- length(players_age$Birth.Year[(players_age$Birth.Year >= 30) & (players_age$Birth
print(n[2])
print("Number of players of Age greater than 50")
n[3] <- length(players_age$Birth.Year[players_age$Birth.Year > 50])
print(n[3])
```

```
[1] "Different Age Range of Players"
[1] "Number of players of Age less than 30"
[1] 246
[1] "Number of players of Age 30-50"
[1] 418
[1] "Number of players of Age greater than 50"
[1] 82
```

```
pie(n, c("<30", "30-50", ">50"), main = "Number of players of each Age range", col = c("p
```



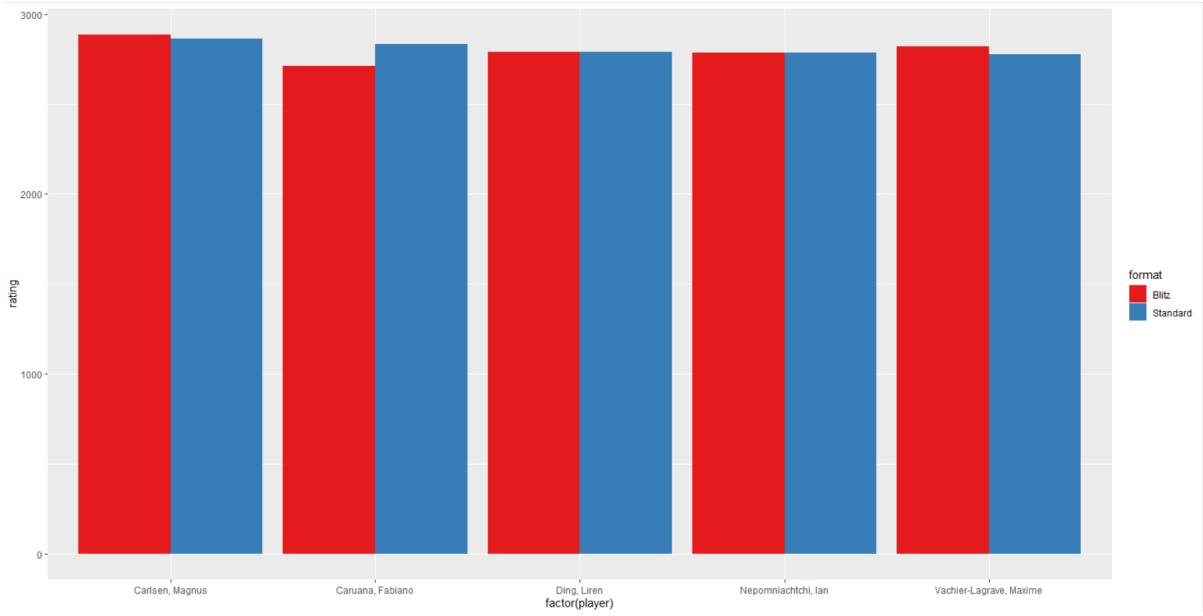
By observing this pie chart, we can notice that most of the Chess Grandmasters are aged between 30 and 50 while there are quite a few younger GM as well who are aged under 30. Old people i.e. aged greater than 50 are less likely to be Chess GMs, quite contrary to the public perception of the sport.

5. Analysis of Standard Vs. Blitz Rating

```
# Comparison Between Standard and Blitz Ratings
players_ratings <- subset(players, select = c("Name", "Standard", "Blitz"))
print("Ratings of the top 5 players")
print(head(players_ratings, 5))
```

```
[1] "Ratings of the top 5 players"
                  Name Standard Blitz
1        Carlsen, Magnus     2863  2886
2       Caruana, Fabiano     2835  2711
3            Ding, Liren     2791  2788
4    Nepomniachtchi, Ian     2784  2785
5 Vachier-Lagrave, Maxime     2778  2822
```

```
rating_gen <- data.frame(
    format <- c("Standard", "Standard", "Standard", "Standard", "Standard", "Blitz", "Bli
    player <- c(
        "Carlsen, Magnus", "Caruana, Fabiano", "Ding, Liren", "Nepomniachtchi, Ian", "Vac
        "Carlsen, Magnus", "Caruana, Fabiano", "Ding, Liren", "Nepomniachtchi, Ian", "Vac
    ),
    rating <-
        c(
            players_ratings$Standard[1], players_ratings$Standard[2],
            players_ratings$Standard[3], players_ratings$Standard[4],
            players_ratings$Standard[5], players_ratings$Blitz[1],
            players_ratings$Blitz[2], players_ratings$Blitz[3],
            players_ratings$Blitz[4], players_ratings$Blitz[5]
        )
)
print(ggplot(rating_gen, aes(factor(player), rating, fill = format)) +
    geom_bar(stat = "identity", position = "dodge") +
    scale_fill_brewer(palette = "Set1"))
```



As we can see in the graph, for the top players, there isn't that great of a difference between their Standard Rating and their Blitz Rating. This implies that time constraint doesn't has that much of an effect on the performance of the top chess players.

Source Code

The source code for this project can be accessed using the following link :

https://github.com/Chaitanya-Raj/Data_Science_Project