ONLINE CHESS GAMES

VISUALIZATION - PROJECT 1

Alena Maria Thomas

30 September 2023

INTRODUCTION

Online chess games have revolutionized the timeless game of strategy, bringing it to a global audience. Players of all skill levels can now engage in exhilarating matches with opponents of varying skills, anytime and anywhere. These games offer accessibility, convenience, and a wealth of learning resources, making chess more approachable for newcomers and a dynamic challenge for seasoned players. Whether you are seeking a quick match or a deep, contemplative battle, online chess provides a rich and diverse gaming experience, uniting enthusiasts and sharpening minds across borders.

Through the analysis of a comprehensive dataset encompassing player ratings, game outcomes, opening moves, and various game attributes, this data analysis project seeks to uncover patterns, and insights within the realm of online chess. By exploring factors influencing player performance such as time controls and chess openings, this project aims to shed light on the strategies employed by players to achieve victory or secure draws.

DATA DESCRIPTION

Link to the dataset used: https://www.kaggle.com/datasets/mysarahmadbhat/online-chess-games/data

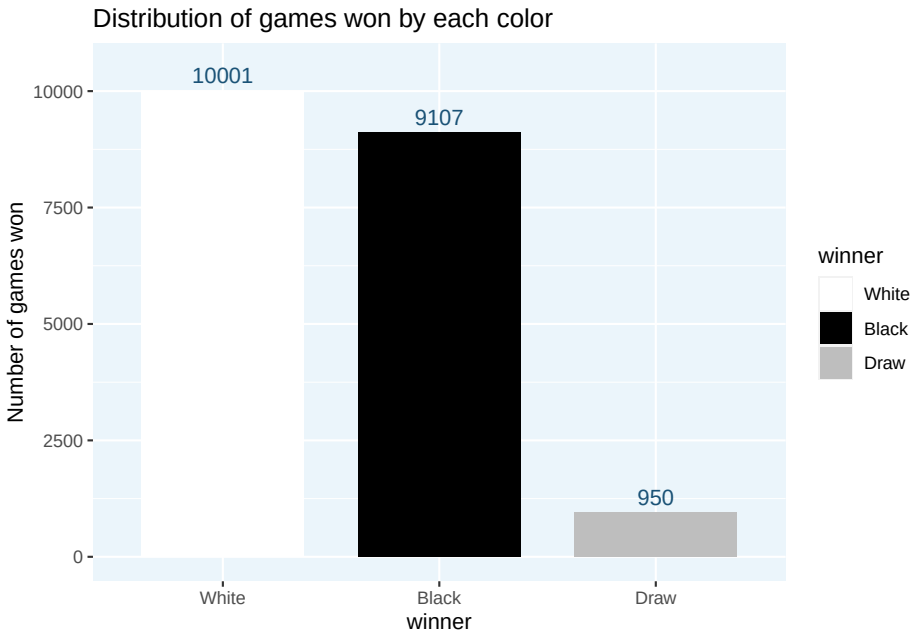
Some variables of interest are:

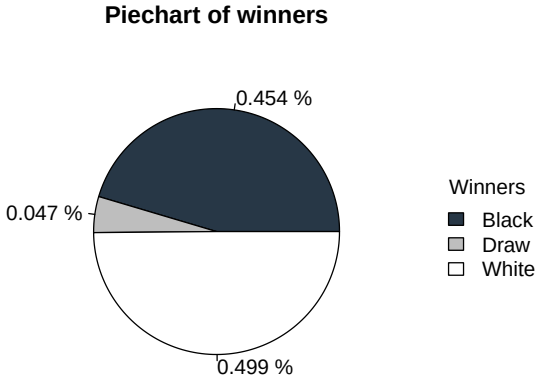
- `turns` (Integer) :The total number of turns taken until the games ends
- `victory_status` (Character): Indicates the status of the game at the end. Possible values are `Draw`, `Mate`, `Out of Time`, and `Resign`.
- `winner` (Character): Indicates the winner of the game. Possible values are `White`, `Black` and `Draw`.
- `time_increment` (Character): The amount of time added to the clock after each move is made. For example in a 15+10 game, you start with 15 minutes and get 10 seconds added to your clock after each move.
- `white_id` (Character): A unique id that identifies the player playing white. The id can consist of letters, numbers, underscores and dashes.
- `white_rating` (Integer): The chess rating of the player playing white. The chess rating is a number that indicates, based on the past performance, how good the you are at playing chess.
- `black_id` (Character): A unique id that identifies the player playing white. The id can consist of letters, numbers, underscores and dashes.
- `black_rating` (Integer): The chess rating of the player playing black.
- `opening_shortname` (Character): The short name of the opening strategy equipped in the game.

EXPLANATORY DATA ANALYSIS (EDA)

Let us explore the dataset by finding answers some questions with the help of visualizations.

What percentage of games were won by white? How many ended in a draw?



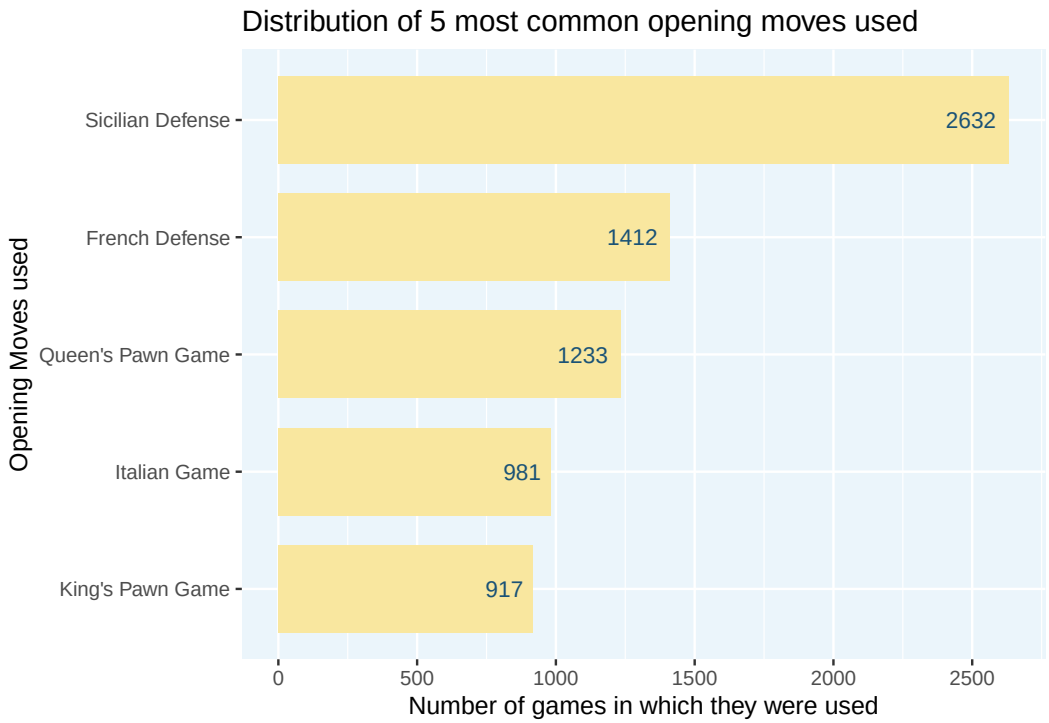


OBSERVATION:

- From the above two graphs, we can observe that white appears to win more number of times as black, though the difference is quite negligible (0.045%).
- Approximately 0.05% of matches ended in draw.

Which opening move was most frequently used?

Since there are large number of different opening moves that have been recorded, we observe the frequency of the 5 most commonly used opening moves.



OBSERVATION: The Sicilian Defense is the most commonly used opening strategy, which is used 2632 / 20058 = 0.131% of the times.

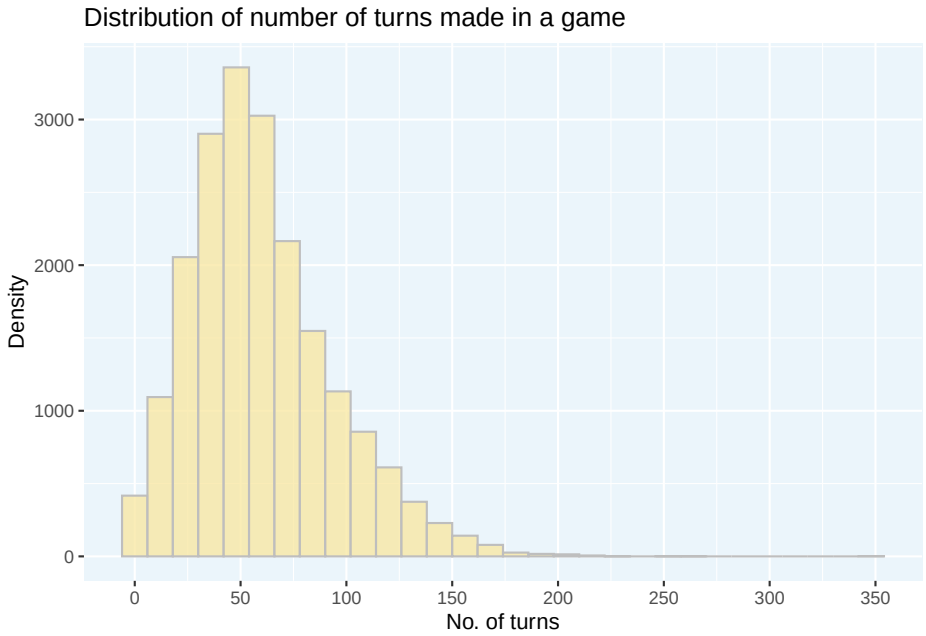
Distribution of number of turns made in a game

We can use the `summary()` function to get information about the distribution of the variable **turns**.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   37.00   55.00   60.47   79.00  349.00
```

This can be visualized with a histogram as follows:

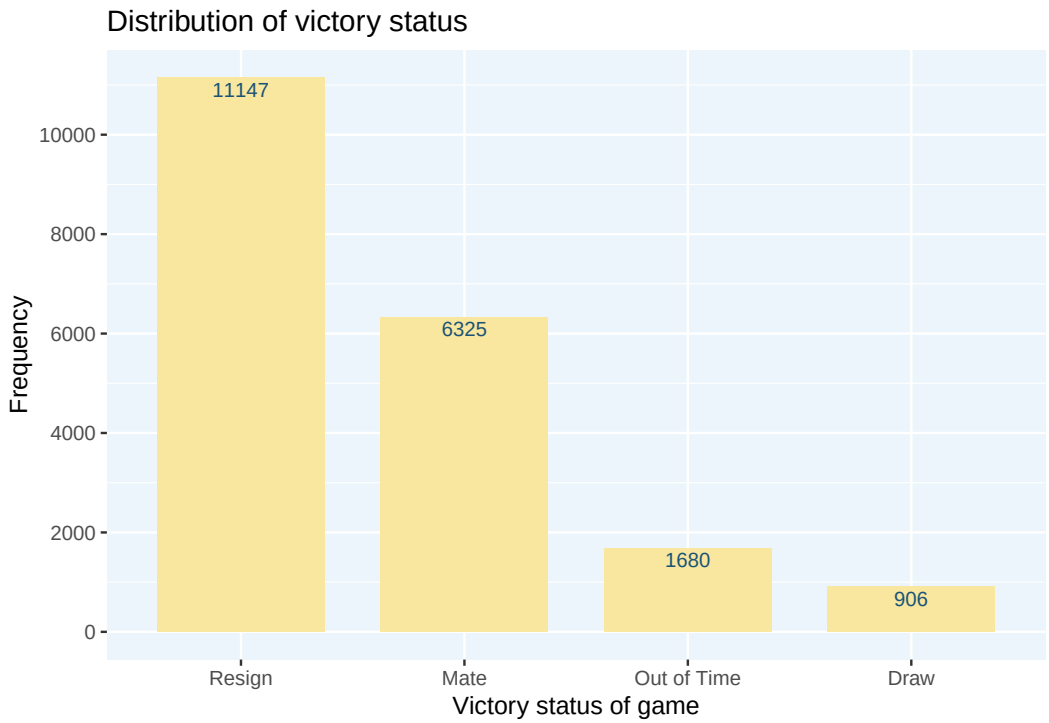
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



OBSERVE: We can observe from the graph the number of turns taken to end a game follows right-skewed distribution, which the mean value approximating 60.

Distribution of the victory status of the games

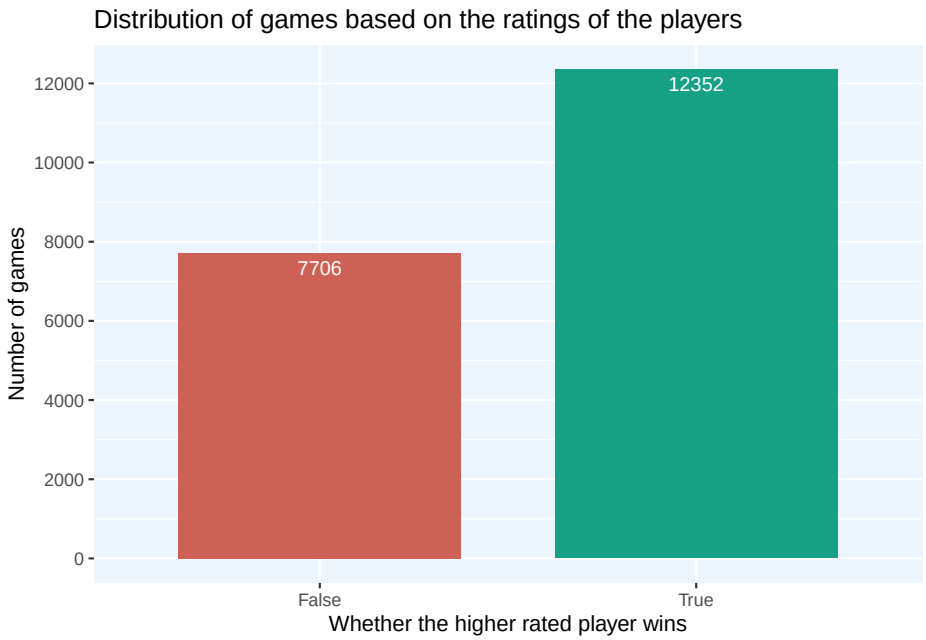
```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```



OBSERVATION: We can observe that the majority of the games end by one of the players resigning the game, which happens approximately 11147 / 20058 = 0.131% of the times.

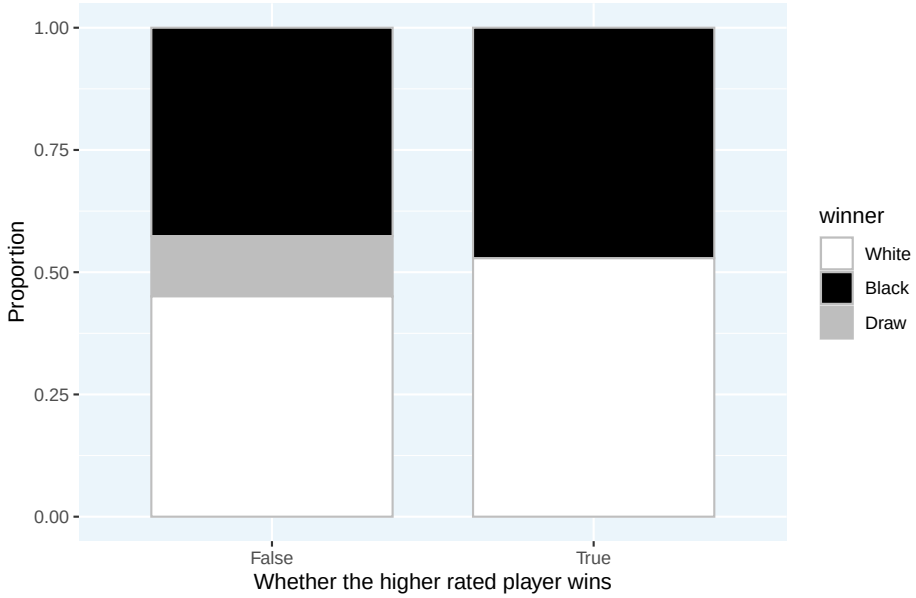
What percentage of games are won by the player with the higher rating? Does this vary by piece color?

```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```



OBSERVATION: We can observe that a lot more games are won by players with higher rating, approximately 12352 / 20058 = 0.131% of the times.

Now, let us explore if this distribution varies with the colour of the player's pieces.



OBSERVATION:

- In games where the higher rated player wins, it is observed that the number of times white wins is very close to the number of times black wins.
- In games where the lower rated players wins, we observe the same thing.
- Hence, we can conclude that the event that the higher rated player wins does not depend on the colour of the pieces of the higher rated player.

SUMMARY

- White appears to win more number of times as black, though the difference is quite negligible (0.045%).
- The Sicilian Defense is the most commonly used opening strategy, which is used 0.13% of the times.
- The graph the number of turns taken to end a game follows right-skewed distribution, which the mean value approximating 60.
- We can observe that the majority of the games end by one of the players resigning the game, which happens approximately 0.55% of the times.
- We can observe that a lot more games are won by players with higher rating, approximately 0.61% of the times.