PATTERNS AND TRENDS IN CHESS GAMES: A DATA BASE ANALYSIS

Adriel Dias, Gabriel Schuenker, Gustavo Luciano, João Vitor Rezende
Affiliation: Fundação Getulio Vargas, Rio de Janeiro, Brazil

Abstract

Games of chess have been a source of research and analysis, for a long time, getting the attention of chess player and mathematicians for centuries, providing dozens of studies about chess, its tournaments and grandmasters. Recently, with the rise of online chess sites, it has become possible to not only get a deeper understanding of high level, tournament play, but also to extract and study the games of novice and average players, providing a better understanding of the differences between the levels of play. The aid of advanced chess engines and artificial intelligence also allows to have a clear understanding of the game's progression and see, numerically, who has the advantage and when it changes. With those tools and a modest database, we aim to find patterns and correlations between a game's result, process and its players by analysing a player's rating, the game's duration and other game factors
Keywords: chess engines; game analysis; data visualisation

I

1 - Introduction

Chess is considered one of the most complex and oldest games that are still played. There are, theoretically, 197,281 different possible positions just for the 4 first moves, and with some games lasting longer than 100 turns, it is reasonable to see how data about a chess game can be very complex and good source for pattern finding. Chess games are the subject of study for centuries, with scholars trying to optimize the game, teaching artificial intelligence to play it, and seeking patterns between the highest-rated player and tournament results. This study also aims to find different patterns at high, low, and intermediate levels of play. Using data collected from online chess games and the stockfish engine information about the players, the general game and every individual move was collected, which was then used to enrich the information and create visualization tools.

2 - Literature review

As stated, chess is not an uncommon subject of study, large scale analysis of tournament play, **[1]** as well as the best ways to evaluate a player's abilities**[2]**, the construction of the best Artificial Intelligence**[3]**. Most of the present literature focuses on higher levels of play, using databases of tournaments and grandmasters, or of high power engines, in this research, we analyze games of different levels, time limits, and turns counting, approximating the average seem on Lichess, and perhaps the average expected games of chess.

3 - Research methodology

The main dataset used for this study was found in Kaggle[5]. The original author provided data about more than 20.000 games collected from the Lichess website. The dataset was then cleaned, to remove games were either too short or duplicated. For a better understanding of the data, quantitative stats like a player's rating or the game's time were turned into qualitative classes.

The clean and added data was then not only used for direct analysis, but also used for game by game, move by move meta-game analysis, providing precise information on the game status by the use of the open source stockfish engine, set with a depth of 10, meaning for every move, we can calculate the average advantage (in pawn pieces).

3.1 - About the Dataset and the research

The original chosen dataset is composed of 16 columns and 2058 rows. It is open source and available on Kaggle. The original author collected the data by checking the game history of different players. Each of the columns represents a different game factor. For this research, the main focus is in the columns "moves", "black_rating" and "white_rating". The first provides a string with every move played, given in standard algebraic chess notation, providing for every movement, the piece moved and the destination of the move. The rating column gave a number, the rating in the glicko[2] system of the respective player.

The dataset provides very rich information. With that, we plan to see how a player's rating relate to the number of resigns, if the game turn count relates to the winner in some way, and how the pawn pieces' used squares relate to a player's rating, and if a players' advantage (evaluated by chess engines) is more variable in lower ratings.

More specifically, the following hypothesis were raised:

- Is an engine's evaluation of a higher rated game less variable?
- Are higher rated players more likely to resign?
- are the lower rating players' pawns movements more disperse across the board?
- Does white hold a higher win rate then black when there are fewer turns?

3.2 Cleaning and treatment of the dataset

The dataset used for the research contained plenty of information, allowing for analysis of numerous topics. Although, for this research's purpose, some changes needed to be made for a better analysis and exploration.

Firstly, it was noticeable that a few games were duplicated. The original dataset was created by collecting every single game in a given player's history, so it is natural to assume that two different players may have played against each other. Using the pandas library in Python, the repeated rows were removed, bringing it down to 19629 rows.

Then, the games with a turn count inferior to three were also removed, as those don't provide enough information. The turn count was chosen for specifically three turns because after the fourth move, the number of possible different board states is more than 190,000, with 8 different ways to end the game, as opposed to the less than 10,000 possible games and no endings for the first 3 turns. It is expected for some games to have some level of asymmetry or even a checkmate after the fourth move.

The columns with the rating of the players, based on the glicko system, were used as a basis to create qualitative levels of play. For our purposes, it was considered that a third of the player base would fit each group, meaning that, for example, high level starts at 1705 glicko, this fits with the provided distribution on Lichess[4]. But may have some problems when considering the difference between the outliers, such as grandmasters professional players.

3.3 - The usage of the StockFish engine

For a better analysis of the moves' column, the Stockfish engine was used, as well as the Python chess library.

The Stockfish engine is a tool able to evaluate the game's advantage based on board states. By analyzing a PGN(portable game notation) of a given chessboard, the engine is able to predict possible following moves by the use of a neural network. The engine is able to work with many different metrics. The chosen metric was "centipawns", which gives a number, with equivalent to 100 being the equivalent of the material value of a pawn piece. Note that the engine may give a player with less material the advantage if it notices advantages in other ways, such as positional, tempo, or even if it calculates a future advantage in a series of plays.

The chess library for Python allows for the games described in the moves' column to be emulated and played by the computer, providing the partial board state each time. Then, the library may feed the board state to the stockfish engine, which allows to see the evaluation of every move, providing a good overview of who is currently more likely to win for each move made.

The Stockfish engine allows for different levels of analysis, with the depth, time for each move, used CPU threads and the size of its transposition table, and other parameters may be changed. The results analyzed for this research were obtained with the use of the games with most parameters unchanged, the chosen depth was 10, and the engine version is Stockfish 16. It should be noticed that changing any of those parameters would slightly change the outcomes and could result in a very long computing time.

4 - Results

This section show the visualisation results for the proposed hypothesis, as well as some brief commentary on the observed values.

4.1 - Expected Results

With the processed dataset and added stockfish information it was expected to see the level of the game be related to the stockfish's evaluation variance, as well as to the number of resigns. For the use of the pawn pieces, it was expected to see the center squares being highly used for all levels of play, but more concentrated in high an medium levels, while low levels would be more likely to move the border pawns. It is expected to see white's win rate being higher in shorter games

4.2 - Observed Results: evaluations

With stockfish's evaluations and the equally distributed levels of play, it was possible to plot a graph comparing the evaluation values of every position seen in every game, divided by the tier of play of the observed player, as shown below:

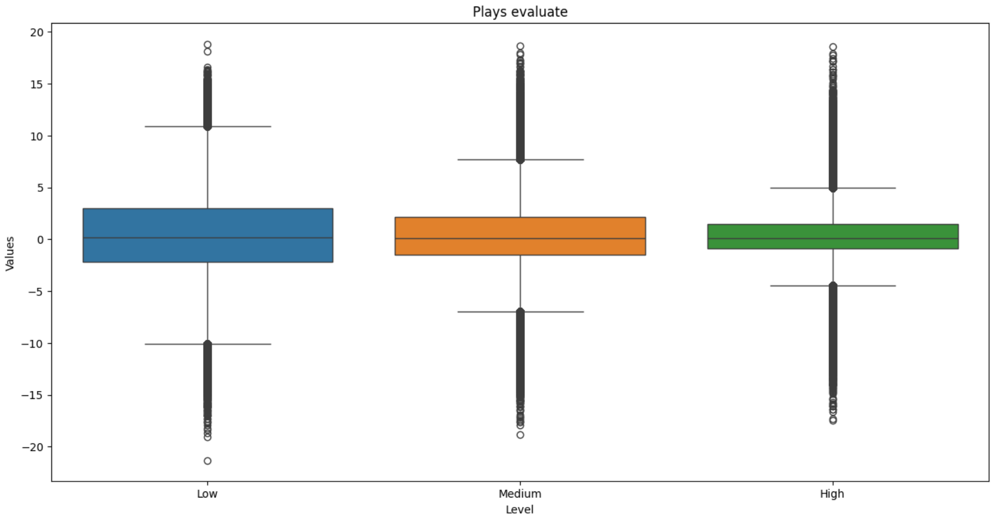

Figure 1: box plot with all of the evaluations by players' level

The graph above indicates a strong relationship between the level of play and the consistence of the evaluation. The lower rating players presents a much higher distance between the lower and upper quartiles, while the high level box plot shows a very small box. The medium level works as a solid middle ground between the two, indicating that this variance really gets lower with higher ratings.

The graph still contains a lot of outliers, and does not provide any information on how quickly the evaluation may change, nor does it show much about the variation on a single game. Figure 2 shows another visualization tool for better understanding: a graph containing the standard deviation observed in Stock fish's evaluation for every game, again, divided between the created tiers of play:

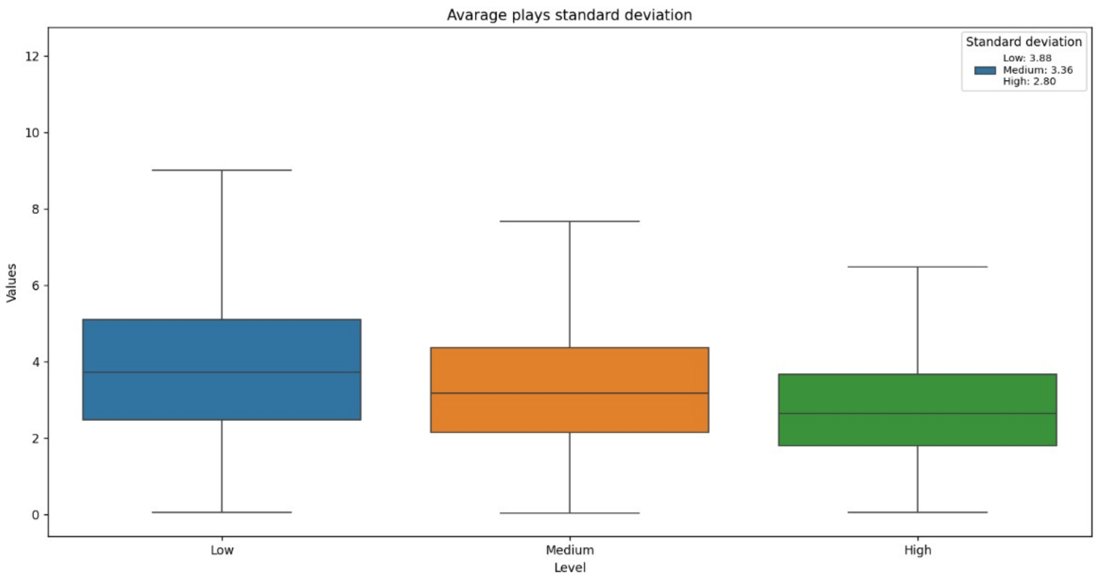

Figure 2: box plot with each game's standard deviation, by players level

Once again, we notice that the boxes become increasingly smaller with higher ratings, not only that, but the maximum value of the deviations also gets smaller, and the indicating more well-behaved games. Finally, it is also noticeable in this graph that the observed values also got closer to zero with the increased player's rating, indicating not only less variance between games, but also within the same game.

4.3 - Observed Results: resigns

To test the level of relationship between a player's rating and the resigning rate, the games were first filtered to show only rated games, as it was expected to have fewer resigns for out-of-game factors this way. A violin graph was the plotted, showing how much of every defeat type happened, in proportion to the defeated player's rating. The following represents the result:

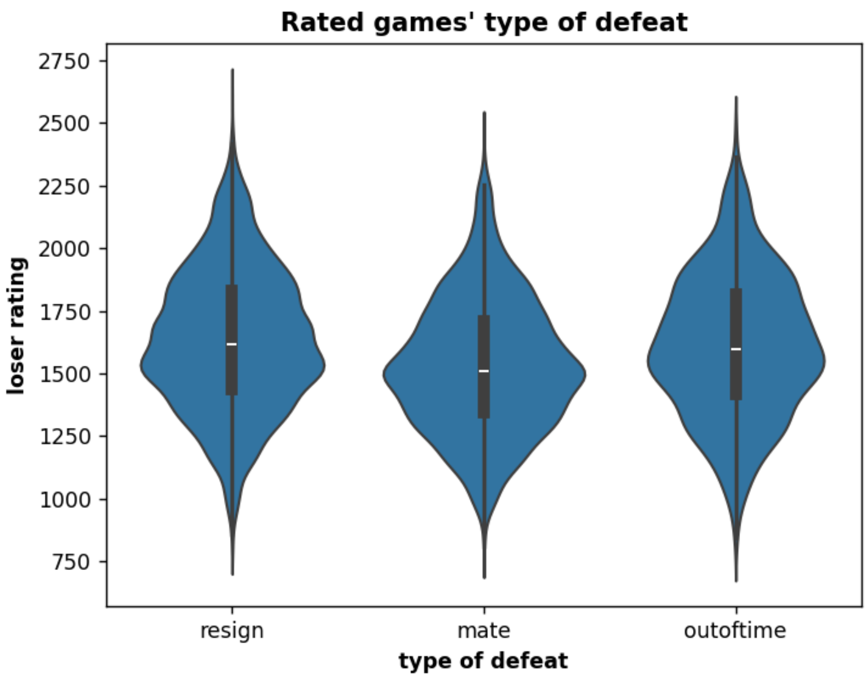

Figure 3: Violin graph of the type of defeat by defeated player's rating

The plotted graphs indicate that, in lower ratings, mate would be the most common way of ending the game, as the out-of-time and resign graphs are considerably thinner in the lower observed rates if compared to the mate's section.

For the higher levels, it can be noticed that the highest rating analyzed only appears on the resign graph, and the highest seem rating for out-of-time ending is also higher than the highest noted value for a mate game end. The rating range between 2000 and 2250 also holds a significantly lower concentration on the mate endings if compared to resign and out-of-time resolutions.

The similarities between the resignation and the out-of-time plotting are also relevant, as it indicates some sort of relationship between those two, but also may impede any strong conclusion for the rate of resignation if compared to all other possible endings.

4.4 - Observed Results: Pawns moves

The pawn moves, represented algebraically as the lack of an uppercase character in the string, were collected and added to a heat map, that describes the number of movements to a given square in a average game

The observed differences may not be as big as initially expected, but we can still see some patterns. In lower levels, black and white used spaces seem almost symmetrical, indicating some level of normality

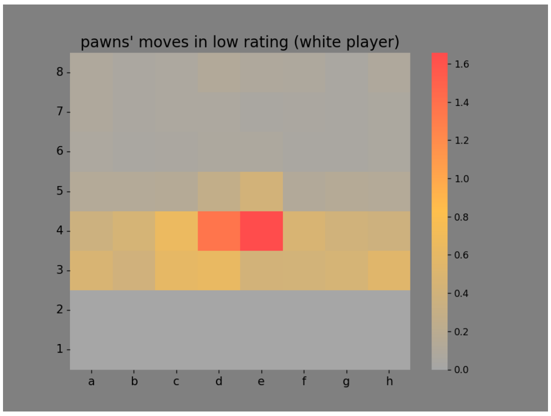

Figure 4: heat map for low rating white pawn players moves

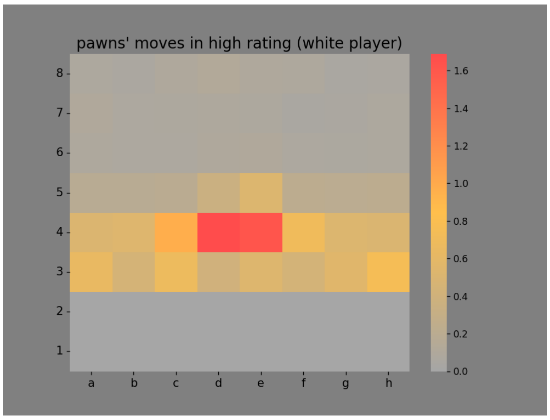

Figure 5: heat map for high rating white pawn players moves

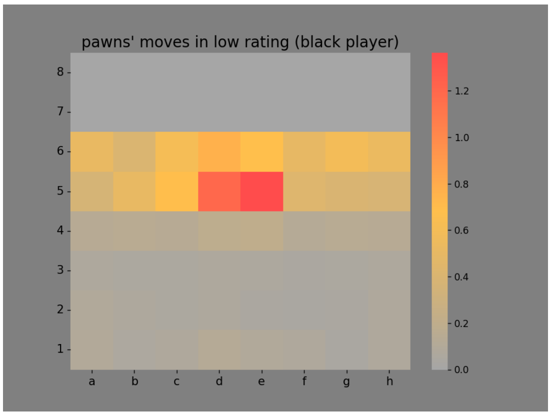

Figure 6: heat map for low rating black pawns players moves

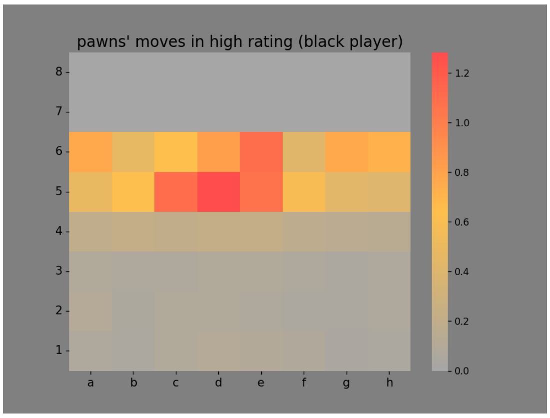

Figure 7: heat map for high rating black players pawns moves

in both the opening and the development of the pieces, with the usage of all of the squares other the d and e files being really low for both players.

In the higher ratings, it could actually be perceived that the use of different squares actually increased on both players' sides, but most notably, the higher-ranking black players seem to heavily use a broad range of different squares, indicating either a broader range of openings or a wider development of the different pawns in a game.

Another difference can be seen by analyzing the files "e" and "d", as the preference for movement of the pawns seems to concentrate in those two for both high and low levels of play. In higher levels, it seems like the most moved squares are d4 for white and d5 for black, with e4 and e5 being close seconds, but in the lower levels, the opposite occurs: e4 and e5 are the squares with the most observed movement, while d4 and d5 seem to be close seconds

4.5 - Observed Results: first move advantage

For the results involving the white win rate based on turn counting, the Dataset was divided in three groups of equal size, making a turn count of less than 43 being considered a "low" count, while games with more than 69 turns fell on the "high" end, and everything in between was treated as "medium", the following image contains the graph plotted with this data, using the seaborn library:

In the plotted graph, it is possible to see a clear, but small advantage of the white player over the

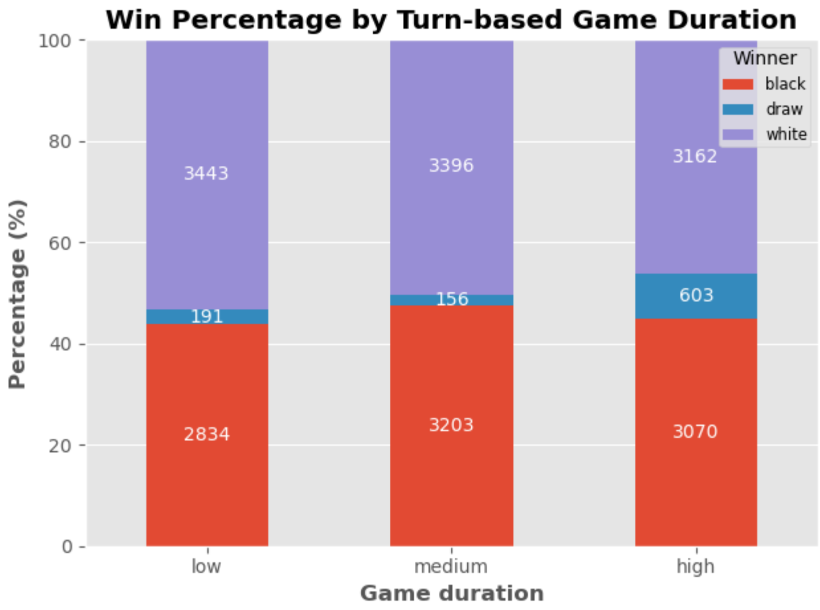

Figure 8: stacked columns graph of winners by turn counting

black player, resulting in a white win rate that starts big and gets progressively smaller with longer turn counting. Notice that ties have highest rate on high turn counting and black actually has the most wins in the medium counting. Thus, it can be observed that, although never too big, the games presented on the Data Set have a margin for more white wins in all the different turn counting levels, but this difference is significantly bigger in the short end of the turn counting.

Discussion

The visualisation was shown above, along with some brief discussion, here, we aim to have some more solid conclusions, as well as state what we can't conclude, by answering the questions raised in the beginning:

- The engine evaluation seems to be way less variable in the higher rankings, indicating consistency as a bigger factor in high-ranking games, rather than the presence of game changing moves. This pattern could also be tied to the other hypothesis, as forfeiting when the advantage becomes apparent might influence the variance in a game.

- In the higher rankings, it is noticeable that there are fewer endings by checkmate, so, proportionally, more games end by resigning, although, the games ended by time out seem to follow the same pattern as the games resigned. It would be safer to conclude that fewer games end in mate in the higher rankings rather than more games are resigned.

- The lower-rated players in the collected data actually are more rigid in their pawns' movements, choosing to move to the same specific squares more frequently, while the higher-rated players seem to use different squares with more frequency, indicating more flexibility. In both tiers, white pawns concentrate in the same squares more than black pawns do, as black usually plays to respond white.

- White's win rates seem to be negatively tied to the games' turn count, especially if compared to black's win rate instead of just of the total of games, analyzing this way we can find numerically that there's a difference of nearly 9 percentile points in the lower turn counting, and of only about one in the higher counting. By looking for the other two outcomes, this relationship seems to no be totally linear, as the black's win rate rises, then shrinks, and the tie rate shrinks, then rises. White's win rate might also follow a nonlinear changing if further analyzed.

Summary

The presented data allows us to already see some patterns clearly. Such as that, between different levels of play, the game becomes less variant, and less games end by checkmate. The open-source nature of the main engines and online chess servers allows the expanding both the number of the games and the accuracy of the analysis. By gathering a greater number of games and more powerful engine configurations, it is expected to have some very reliable and more conclusive data. The addition of grandmasters games and the comparison between professional players and the previously selected groups could also enrich this analysis, allowing to see in which points they diverge from other high-rated players, in the same way that the gathered data could separate the high-rated from the low-rated. It is expected that the enrichment of this research would be able to form an even more conclusive analysis, as well as to allow a broader scope of game and even meta-game analysis.

References

[1] Mathieu Acher and François Esnault, *Large-scale Analysis of Chess Games with Chess Engines: A Preliminary Report*, Research Centre Rennes – Bretagne Atlantique, 2016.

[2] Mark E. Glickman, *The Glicko system*, *Applied Statistics*, 1999. Available at: http://www.glicko.net/research/glicko.pdf.

[3] Q. A. Sadmine and A. Husna, *Stockfish or Leela Chess Zero? A Comparison Against Endgame Tablebases*, in *Advances in Computer Games*, 2024.

[4] *Lichess Rating Distribution for Blitz*, accessed 2024-10-10. Available at: https://lichess.org/stat/rating/distribution/blitz.

[5] *Chess Game Dataset (Lichess)* , accessed 2024-11-10. Available at: https://www.kaggle.com/datasets/datasnaek/chess