# Chess Analysis Project 1

Mike Birlew

2023-02-11

*Introduction*

Chess is a complex game of strategy with several factors involved that lead to a win. The raw data set comes from kaggle's website https://www.kaggle.com/datasets/datasnaek/chess. This data set includes over 20000 games that were collected from an online chess simulator containing user data from Lichess.org. This project will begin to look into patterns that players may use to influence there chances of winning specifically move sets, known as opening moves, that have been defined as standard opening plays consisting of a specific set of moves.

*Data Cleaning*

To analyze the data for move strategy comparison, the raw data has been cleaned from sixteen columns to seven columns. The winner column has been filtered of all games that are not defined as a win leaving 6,325 total games. The standardized opening moves sets (opening_eco) has been factored to contain 272 different opening moves.

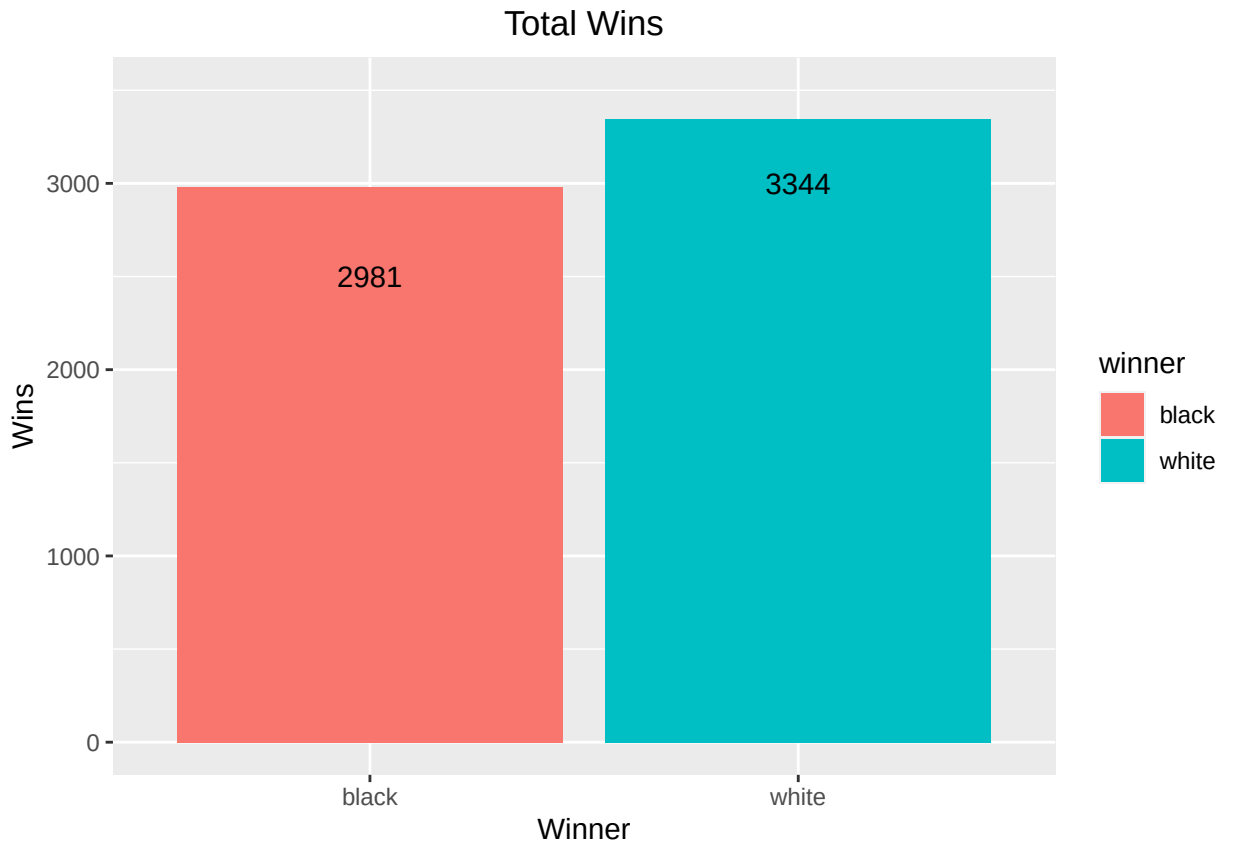
Table 1: Table 1. Chess Data Annotation

| Parameter | Data_type | Discription |
|---|---|---|
| rated | number | Games are rated or unrated in play |
| turns | integer | The amount of turns for a game |
| winner | character | Winner of the match by color |
| opening_eco | factor | Opening move assigned 1-272 |
| opening_name | character | Name of opening move used |
| opening_ply | integer | Number of moves in the opening phase |
| winner2 | number | Dummy variable assigned to winner color (1=White, 0=Black) |
| game_type | number | Dummy variable assinged to type of game (1=Timed, 0=Un-timed) |

(Table 1.) the data set being utilized contains three numeric parameters which include a 'rated' parameter that determines whether a game is added to a persons total game score, the 'game type' that is used to signify whether a game was used under Fischer rules, a typed of timed play and a 'winner2' dummy variable assigned to which color won the game. The two integer parameters are 'turns' indicating the number of total turns played in the game and an 'opening_ply' which indicates how many moves were included in each opening move set. The character variables are 'winner' displaying which color won the game and the 'opening name' which is the name of the opening moved used to begin the game. The 'opening echo' is a factored parameter for each type of moves set.

*Data Analysis*

When looking for advantages in play fig.1 shows that 'white' wins more often than 'black'. In the game of chess 'white' always moves first. This is a well studied phenomenon in chess and the consensus that follows is called 'First-move advantage' https://en.wikipedia.org/wiki/First-move_advantage_in_chess It generally follows that with the first move 'black' is reacting to white as the game continues. Out of 6,325 games 'white' wins 3,344 of them. This 53% advantage is consistent with other studies that have been done on 'First-move Advantage'.




(Figure.1) The total number of games in the data set (6,325) divided into wins between 'white' and 'black' shows white winning a 53% of games in what is called 'First-move Advantage'.

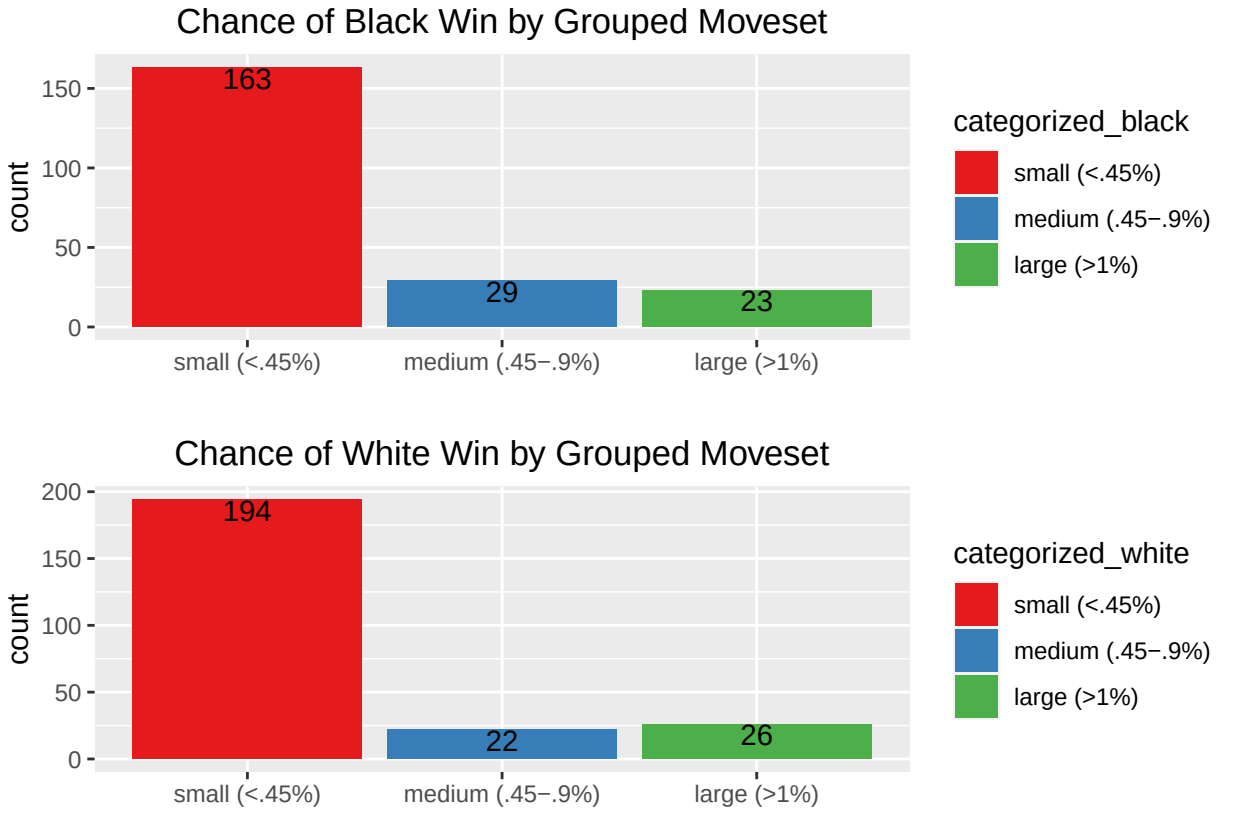
Outside of First-move advantage, are there any other factors that influence the game, especially any that 'black' can use to overcome 'white' always moving first and having First-move advantage? Table 2 shows a summary of the total wins and percentage of wins per each move set. There are some interesting numbers that come out of this table for 'Black Wins'. For each move set, there is a set of moves that have a larger advantage of winning the game despite 'black' winning a slightly less percentage of the games overall. 'Black Wins' shows a larger median and mean when it comes to specific move sets winning games against 'white'. There is also a particular move set that, when used, can increase 'black' wins. This is seen at the 'Black Wins' Maximum value with a total of 281 wins at 9.4% against 'White Wins' Maximum win move of only 171 at 5.1% of wins. When normalized against all games played (6,325), the Maximum wins of the top two move sets is 'black' at 4.4% and 'white' at 2.7% for a 1.7% advantage over 'white' when both players are using the most winning move set.

**Table 2. White/Black Win & Percent to Move Sets**

| | Black Wins | Percent to Win | | White Wins | Percent to Win |
|---|---|---|---|---|---|
| | Min. : 1.00 | Min. :0.03355 | | Min. : 1.00 | Min. :0.02990 |
| | 1st Qu.: 1.00 | 1st Qu.:0.03355 | | 1st Qu.: 1.00 | 1st Qu.:0.02990 |
| | Median : 4.00 | Median :0.13418 | | Median : 3.00 | Median :0.08971 |
| | Mean : 13.87 | Mean :0.46512 | | Mean : 13.82 | Mean :0.41322 |
| | 3rd Qu.: 12.00 | 3rd Qu.:0.40255 | | 3rd Qu.: 11.00 | 3rd Qu.:0.32895 |
| | Max. :281.00 | Max. :9.42637 | | Max. :171.00 | Max. :5.11364 |

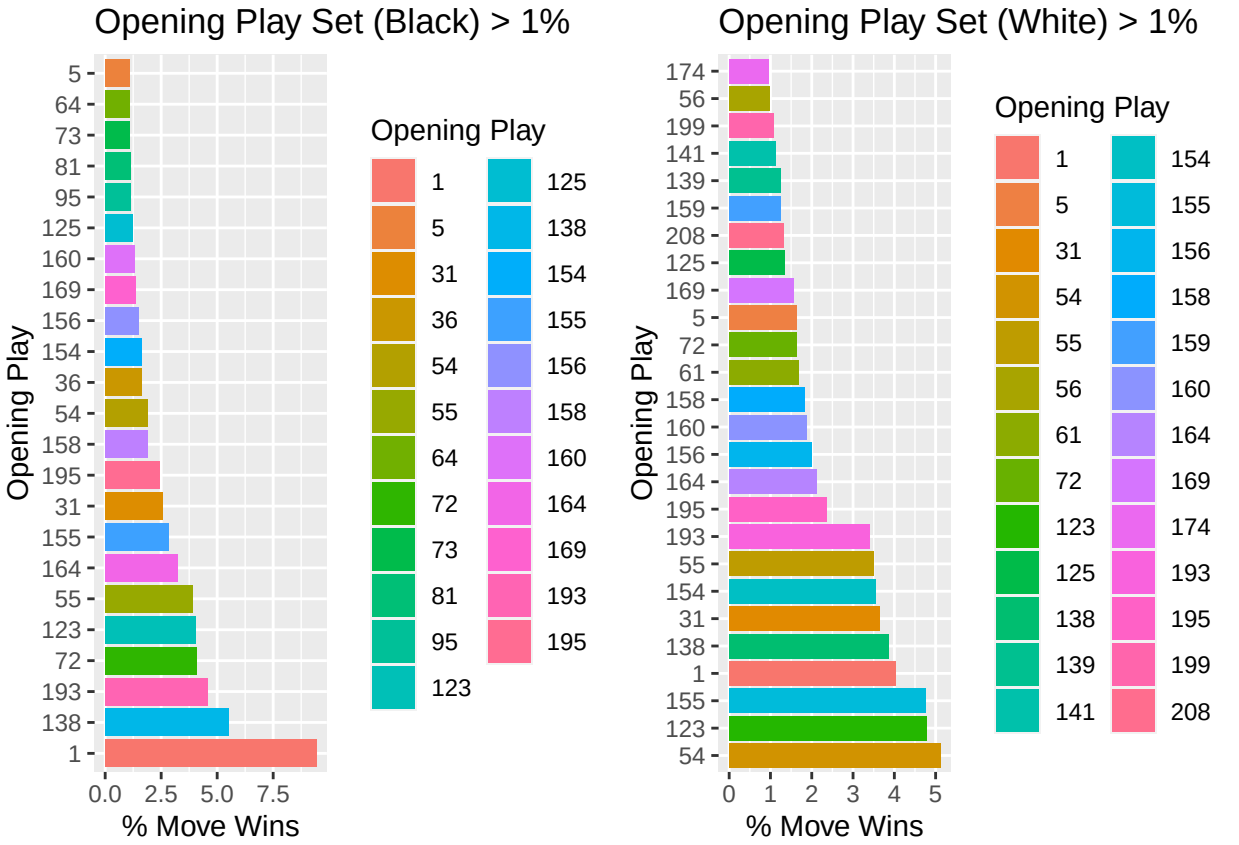
(Table 2.) Shows a comparison of total moves to wins against 'black' and 'white', particularly intersting is the opening move set that increases wins when used by 'black' to 9.4% . When normalized against 'white' wins the max win percent is still higher at 8.4% compared to 'white' max percent win at 5.1% giving 'black' a 3.3% advantage over the strongest move for 'white.

Using the information from Table 1., the data can further be categorized into the most successful move sets for winning based on the total wins of each color. Using the combined mean of both 'white' and 'black' the categories are separated into chances of win at Small (<.45%), Medium (.45 - .9%) and Large (>1%). Figure 1. Shows out of all the wins for a particular color, how many move sets have the highest chance of winning a match. 'White' shows a higher count for 'small' and 'large' chances of winning between move sets but interestingly 'black' has move sets that give a better medium advantage within the categories. If there is a possibility of overcoming the ~3% First-move advantage that 'white' has, it may lie in these opening moves.




(Fig.1) Shows a categorized approach of looking at the move sets that have the highest chance of influencing wins out of all wins within the group. Out of the 272 different move sets listed, white has a total of 48 that are above .45% win ratio with black having 52 above .45% win ratio.

Exploring 'white' and 'black' win ratios further and looking into identifying all top move sets greater than 1% can show the top three winning moves. Figure 1. shows two moves in particular for 'black' that out performs the rest of the opening plays. Move 1 refers to A00 Polish (Sokolsky) opening which consists of over 9% of all wins by 'black' it is interesting to note that 'White also uses this move with a 4% win ratio. The second opening move set that has the highest win ratio is also 'black' with 138 representing C20 Kings pawn game at 5.5%. Kigs pawn is also used by 'white' at an ~ 3.8% win ration. 'White' has the third highest win ration with opening move 54 which is Owens Defense. This move is also used by 'black' in the top 1% of opening moves but is not as successful falling under a 2.4% win ratio.




(Fig.2) Shows the list of move sets in the order of top performers greater than 1% for 'black' and 'white'. Move set #1, A00 Polish (Sokolsky) opening is a clear opening to learn for black to battle 'white' First-move advantage.

*Conclusion*

This is just a small look into certain advantages that a chess player can learn about when trying to better there playing strategies. To see such a stark difference in advantage between opening move sets could bring on much more understanding in how these opening moves could play against one another. One thing of interest is within the top three move sets there are also several variations that can be played to achieve the opening move, there is more than one way to Rome. It would be interesting to create subsections of these opening moves and then compare their efficiency against other opening move sets. The last figure also shows that many of the top move sets used are used by both 'black' and 'white' but with varying degrees of success. One thing is clear though and that is it would benefit any player to learn these top three moves and their variations to gain a substantial gain in their game play.

**Appendix**


1. "365Chess.com" 2007-2023 https://www.365chess.com/eco.php. Accessed 22 February 2023
2. Wikipedia https://en.wikipedia.org/wiki/First-move_advantage_in_chess Accessed 20 February 2023
3. [Github code]https://github.com/mbirlew/ChessPack/blob/main/docs/chess_analysis.Rmd