

Project Milestone 3 – Group 17

[Dataset Link](#), [GitHub Link](#)

Introduction

Our project focused on creating an interactive dashboard for chess players of all skill levels. We focused on answering questions around identifying how a player is doing compared to other players in the dataset. More specifically we aimed to answer the questions: What are the most common openings played? Which openings give White or Black the best chance of winning? What are the most common ways games end (e.g., checkmate, resignation, draw)? What are the most common responses to specific moves? How do time controls affect game length and outcomes?

To support this dashboard we used a dataset from Kaggle including over 20000 games. The dataset has a variety of columns to use for our dashboard, including if the game was rated or not, the number of turns, both players' ratings, opening name, all moves played, the outcome and method of winning. This information allows us to answer the questions about player strategy and performance.

Literature Review

Websites such as Chess.com or Chess Monitor that have dashboards showing a breakdown of games by time control and show statistics such as wins, losses, draws and number of games played. These dashboards only show trends for an individual. This requires a player to have their own chess data to import or have an account to be able to identify their strengths and weaknesses. As a result, it becomes difficult to identify general patterns across the broader chess community.

Other dashboards include heatmaps, decision trees and graph networks that analyze the current position to visualize future positions. These tools allow users to select moves and identify possible counterplay. However, these methods can be computationally expensive, especially in the middlegame and endgame when the number of possible moves increases significantly. To address this, filtering or pruning based on chess engine evaluations can be used to focus on relevant positions.

We used the packages ggplot2, plotly, ggiraph, heatmaply and shiny to plot and add interactivity to each graph. These allow users to explore trends in thousands of chess games with informative visualizations following best practices in data visualization.

Visualizations

The shiny application for analyzing chess game data includes a series of visualizations designed to address specific questions about the games, players, and strategies. The application provides a user-friendly interface with multiple filtering options, such as selecting a

winner, victory status, opening name, and the ratings of both players. These filters allow users to explore various subsets of the data, which are reflected in the interactive visualizations.

The "Win Counts by Opening" chart displays how various chess openings perform across different games, with the data breaking down win rates for both players. This visualization answers the question of which openings tend to result in more wins, making it easier to compare the relative success of various strategies. By using a stacked bar chart, it allows for clear differentiation of win rates between different openings, with the percentage values inside the bars making the data even more accessible. This is useful for players looking to analyze which openings may yield higher chances of success based on historical data.

However, while the chart does a great job of categorizing and providing win rates for different openings, it leaves unanswered questions regarding the nuances of how openings perform against specific opponents. For example, the chart doesn't provide insight into whether particular openings perform better in certain time controls. Additionally, while the chart includes percentages inside the bars, there could be more explanatory context, especially for less experienced players who may not be familiar with the significance of certain openings. Improvements could include adding a descriptive tooltips or legend to help explain what a certain opening does and why it is significant.

Chess Data Explorer

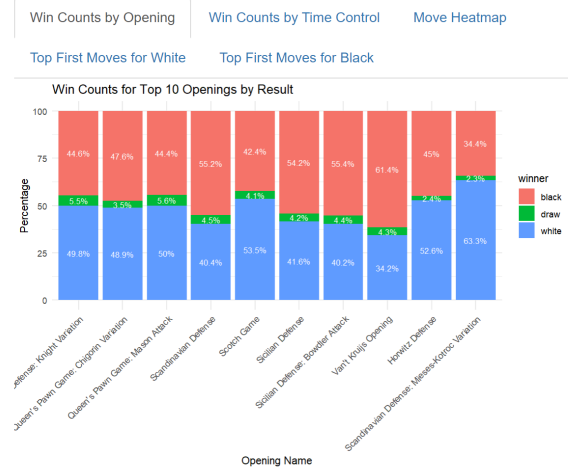
Winner:
All

Victory Status:
All

Opening Name:
All

White Player Rating:
784 1,404 2,700

Black Player Rating:
789 1,409 2,723



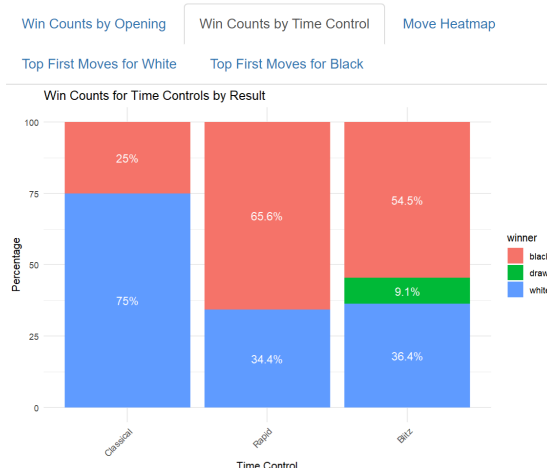
Winner:
All

Victory Status:
All

Opening Name:
King's Pawn Game: Leonhardt Variation

White Player Rating:
784 1,404 2,700

Black Player Rating:
789 1,409 2,723

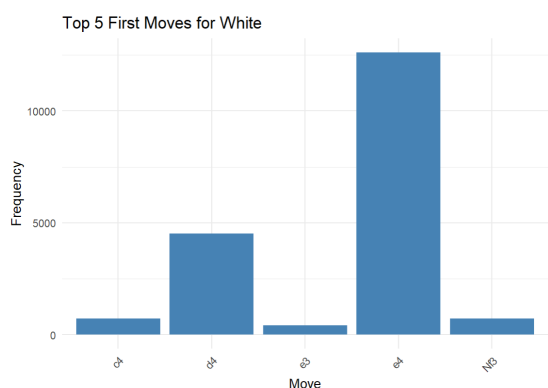
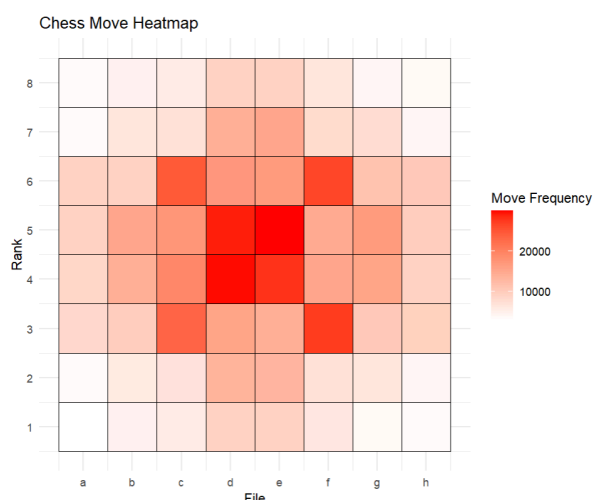


The "Win Counts by Time Control" chart explores the impact of time control on the outcome of games. It answers questions of how different time formats affect the win rates of games. The chart clearly shows how time pressure influences the results, with distinct differences between fast-paced formats (Bullet and Blitz) and slower formats (Classical). This is useful for understanding the dynamics of faster games, where the players often rely more on intuition and speed, versus slower games, where strategic thinking is prioritized.

While the chart effectively contrasts win rates across different time controls, it doesn't provide insights into player skill levels or the effectiveness of time controls for different rating groups. For instance, it doesn't address whether higher-rated players perform better in faster or slower time controls, or how time control influences mistakes or blunders. Adding a breakdown of win counts by player rating or performance metrics across time controls could enhance its utility. To improve this chart, adding clearer labels defining each time control (such as typical time limits for Bullet, Blitz, and rapid) could help provide context to users unfamiliar with them.

The "Move Heatmap" visualizes the frequency of moves made to each square on the chessboard, showing the most contested areas of the board during the games. This visualization answers the question of which areas of the board are most frequently targeted by players, offering a valuable look into how pieces are moved during a game. The color gradient, with red indicating the highest frequency of moves, makes it easy to identify areas of heavy action. This is particularly useful for analyzing whether players focus on central squares or expand their movement to other parts of the board.

While the heatmap effectively highlights which squares receive the most attention, it doesn't answer questions about how specific opening strategies or time controls influence player movement across the board. For instance, the heatmap does not show if certain opening moves lead to more activity in specific regions of the board or whether faster-paced games result in more or less movement in key areas. To improve the heatmap, adding more filters, for example one for time control, could help provide a more comprehensive view of how different variables influence the frequency of moves. Also, allowing users to toggle between perspectives, such as White's and Black's view of the board, could also enhance its accessibility.



The "Top First Moves for White and Black" charts present the most common first moves for both players at the start of the game, answering the question of which openings are most frequently played at the start of a chess game. These visualizations are helpful for both novice players who want to understand common opening strategies and experienced players looking to refine their approach. By ranking the most popular first moves for both White and Black, it provides quick insights into what strategies dominate the early game phase.

However, the chart leaves some important questions unanswered, such as how the first moves correlate with win rates or whether certain first moves are more successful based on player skill level. For example, it would be valuable to see not only which moves are most popular but also which ones are most successful. The chart also doesn't account for variations in first moves depending on time control or rating groups, which could offer more tailored insights. One way to improve the chart would be to add win percentages or success rates for each first move, allowing users to understand not just popularity but also effectiveness. Incorporating player rating ranges or time control filters would add another layer of depth, revealing how different player groups approach the opening phase of the game. Providing explanations or context for each first move could also make the chart more accessible for less experienced players, offering them a better understanding of the strategic reasons behind certain moves.

These visualizations provide valuable insights into different aspects of chess gameplay, but they could be further enhanced by adding additional layers of context, such as player ratings, success rates, and time controls. By incorporating more detailed filtering options and providing explanatory context, these visualizations could become even more useful tools for both casual players and serious analysts.

Conclusion

Our interactive dashboard serves as a valuable analytical tool for chess players seeking to enhance their performance through data-driven insights. While many players can reach a rating between 1500 and 2000 without formal study, progressing beyond this range typically requires a deeper understanding of chess theory. One of the most effective ways to gain a competitive edge is by studying common patterns in gameplay. Our visualizations aim to make this kind of strategic preparation more accessible.

Using our dashboard, we identified that some of the most commonly played openings include the following:

- Ruy Lopez (Spanish Game)
- Sicilian Defense
- Queen's Gambit
- Scotch Game
- Scandinavian Defense
- Van't Kruijs Defense
- Queen's Pawn Game

These openings are widely used at all skill levels, suggesting that these are appropriate openings for all types of players. Notably, we found that the Scandinavian Defense - Kotroc variation and the Horwitz Defense yielded particularly strong results for White, while the Van't Kruijs Defense and the Sicilian Defense were among the most successful options for Black. These patterns can help players to choose openings with favorable statistical outcomes or prepare for likely responses from their opponents.

In addition to openings, we explored how time controls impact game results. Our “Win Counts by Time Control” chart revealed clear distinctions between faster formats such as Bullet and Blitz and slower ones like Rapid or Classical. Players relying on intuition may perform better in short formats, while deeper strategic planning is more prominent in longer games. Furthermore, by adjusting the “Victory Status” filter to focus on games lost from running out of time, we found that time management is a critical factor. White players frequently lost on time when playing the Van’t Kruijs Defense, while Black players encountered similar issues in the French Defense - Knight Variation. This suggests that some openings may be more time-consuming to execute effectively, and players using them should be especially mindful of their clock.

One of the most visually compelling visualizations in our dashboard is the “Move Heatmap,” which highlights areas of the board that receive the most activity. Central squares consistently showed the highest density of moves, reinforcing the classic principle of chess strategy of controlling the center. Positional play centered around these squares often leads to better board control, which can dictate the flow and outcome of the game. This heatmap offers a quick but powerful reminder for players to focus on such areas during both the opening and mid-game.

Altogether, our dashboard not only reveals major patterns in chess gameplay but also translates them into strategies for improvement. With further enhancements such as deeper move analysis, an interactive heatmap suggesting common responses to a certain move, or time usage analysis across ratings, this tool has the potential to support players of all levels in making more strategic decisions in their chess games.