

# Chess Openings for Early Advantage

Lukas Motykowski, Thomas Leisure, Cameron Brandt; Advisor: Dr. William Eberle

## Project Goals

Chess is a battle from the first move to the last. Every opening in chess is unique, and often end in vastly different positions. A well played opening can turn an early lead into a complete victory.

Finding out what opening you should learn is quite an intimidating task due to the sheer volume of different openings. We set out to make a list of openings that show an increased percentage of wins for both white and black.

## Methods and Tools

- Our primary tool was the programming language R
- We used ggplot2 for data visualization, dplyr and plyr for data manipulation, and python for converting PGN files to csv
- Our Dataset came from the public Lichess database of chess games played on their website in January 2021. This consisted of 200GB worth of games that we condensed into 900,000 games total.
- In order to accurately observe win rate, we grouped the games into bins by ELO rating. We then calculated the total wins by black, wins by white, and draws for each opening code. Using this we calculated the win rate of each opening and analyzed them using graphs while also considering the amount of times an opening was played. Openings with a high win rate and more than a few games in the dataset were considered.

## Recommendations and Future plans

- For those who wish to look into this further I would highly suggest using an engine like stockfish to evaluate the position of each of the openings. The engine could also be used to count the amount of blunders made by an opponent. There might be a correlation between certain openings and the amount of blunders played by white or black. If an opening is more prone to blunders it would definitely be one to avoid

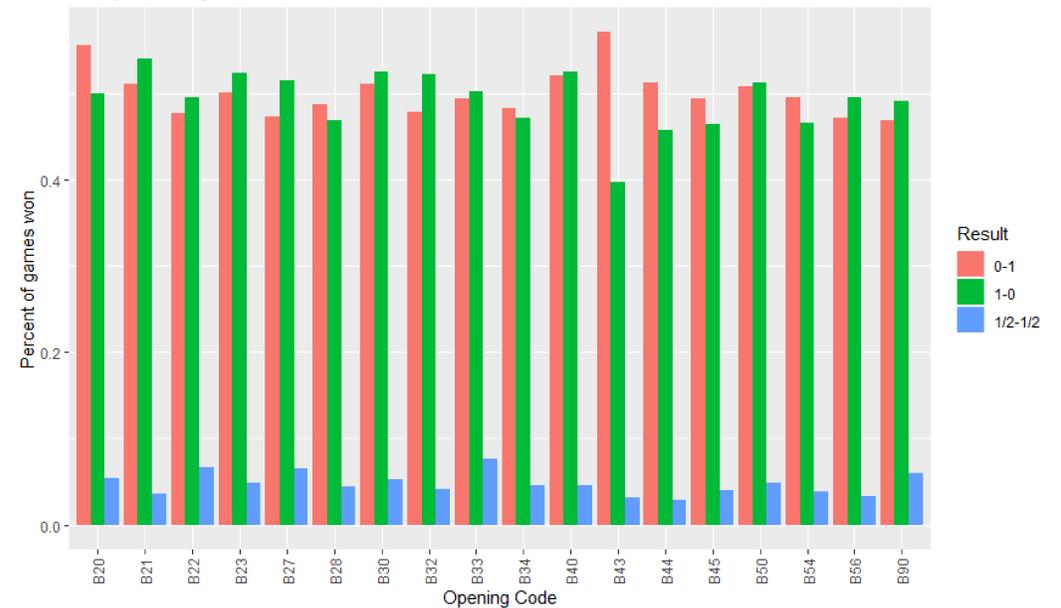
## Results

- While looking at the win rate of each opening we found that for the vast majority of openings, win rate was fairly close to 50% win 50% loss. There were a few outliers that seemed to have a fairly high percentage of wins comparatively.
- For black, Sicilian Defense showed a few variations that seemed to suggest a nice win rate. Most notably the Sicilian Kan showed the highest win rate for the Sicilian variations. The Kings Gambit declined also showed a large win rate for black. Lastly the Caro Kann defence showed a high win rate for black.
- When it comes to openings for white, it almost seems like the vast majority of openings gave white an advantage. A few outliers include the Kings Gambit accepted, King's Indian, and The English opening,
- The rates at which white wins vs black winning seem to suggest that white has an inherit advantage regardless of opening played. This could be due to the fact that white makes the first move, and is thus allowed to take the center in the beginning.

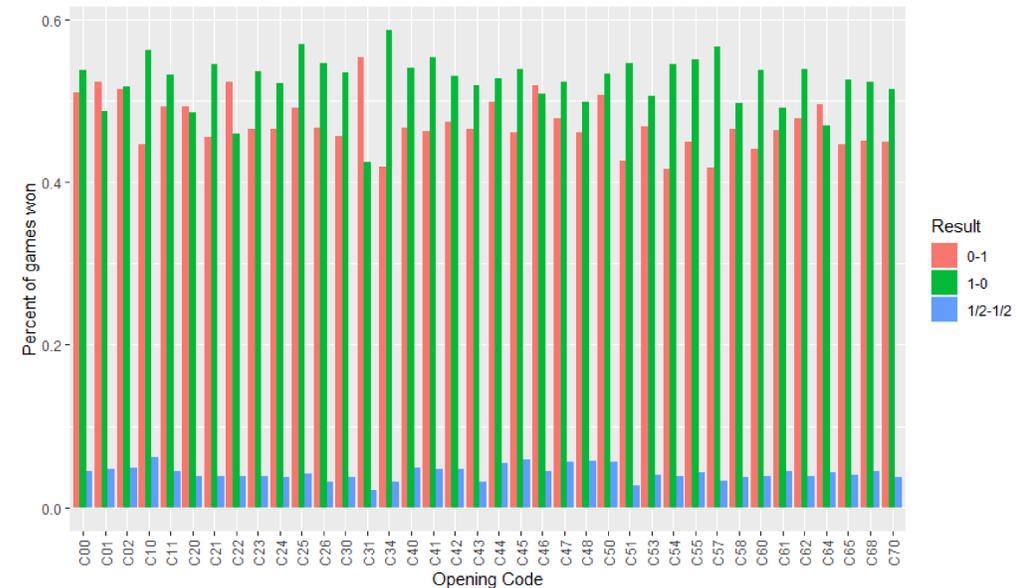
## Conclusions and Discussions

- Although there were some trends in the graphs we developed, we came to the conclusion that although the opening is fairly important there is far more that goes into a chess game than the first few moves. An opening can set up a position but if the player blunders it doesn't matter how good the position is. On the other hand, our data seems to suggest that white will generally win more often than black. This makes the openings in which black had a considerably higher win percentage of particular interest. We think that the cases where black has a much higher win percentage implies that the opening played leads to a weak position for white.
- Based on the in win percentages, our data also seems to suggest that openings that start with e4 end with white winning more often than black.

Win percentage of Sicilian Variations



Win percentage of Different e4 openings



Win percentage of Different d4 openings

