MPCS 56430 - Scientific Computing
Final Project: Chess Database Analysis
Executive Summary

Shane McVeigh smcveigh

1 Abstract

For my project, I applied the high-performance computing and multiprocessing principles covered in the course to a chess game database hosted on Lichess, the most popular open-source online chess website whose game database holds about 30 GB of data monthly. I sought to answer temporal and qualitative trends about the characteristics of chess games. Through effectively pipelining the data using python's `Zstandard` and `multiprocessing` libraries, I was able to effectively gather data and answer my initial questions. I found that high-level players have about 3% higher win rate with white pieces than lower level players, that the proportion of short time-control games has increased over time, while long time-control has decreased, and that there are trends in opening popularities over time and also average player skill level by opening. Full details are in the main project write-up.