CS408X

Final Project - Process Book

Team

; DROP TABLE teams ;

Names

Matthew Fortmeyer, Kien Nhan, Akash Shaji, Jimmy Tran

Overview and Motivation

The purpose of this visualization is to provide an overview of strategy by the top chess players around the world. The goal is to be able to provide clear indications on which moves are “smarter” to make based on what professional chess players have done in the past. The motivation behind this project is that the entire group is terrible at chess and wanted a tool that could give an upper-hand in learning new strategies or moves that would give a higher win percentage overall.

Related Work

https://lhartikk.github.io/chessopenings/

http://www.chesstree.net/

Questions What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

When the project was initially considered, there were a few questions that an answer was desired.

Based on the history of professional chess players and their recorded games...

- What is the most effective strategy for playing chess to win?
- How can that be figured out through data?
- What moves are the most common as starters?
- What starting moves are most to the player?

Each of these questions are the basis for each of the visualizations and data analysis for the project.

Data Source, scraping method, cleanup, etc.

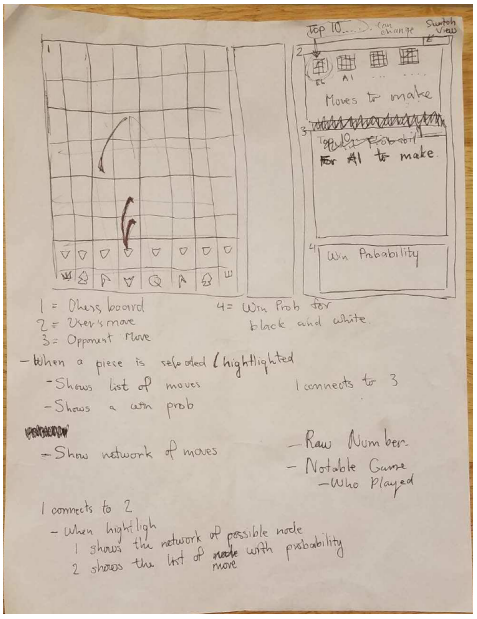
The data was taken from chessgames.com. A python script was used in order to scrape the data from the website since chessgames.com does not have an API. The script then iterates through each of the player's games through their profiles and pull the text. During this process a system accidently got permanently banned from the website due to scraping too much data too quickly. While chessgames.com provides the games in PGN format, which is one of the standard formats for storing chess games, it did so rather inconsistently so we had to deal with a lot of cleanup (one of the more interesting cases is where someone died before the game finished so it just had a * for the result). Afterwards, the PGN was parsed and stored into the JSON format, separating moves in order to generate a graph representing the list of moves.

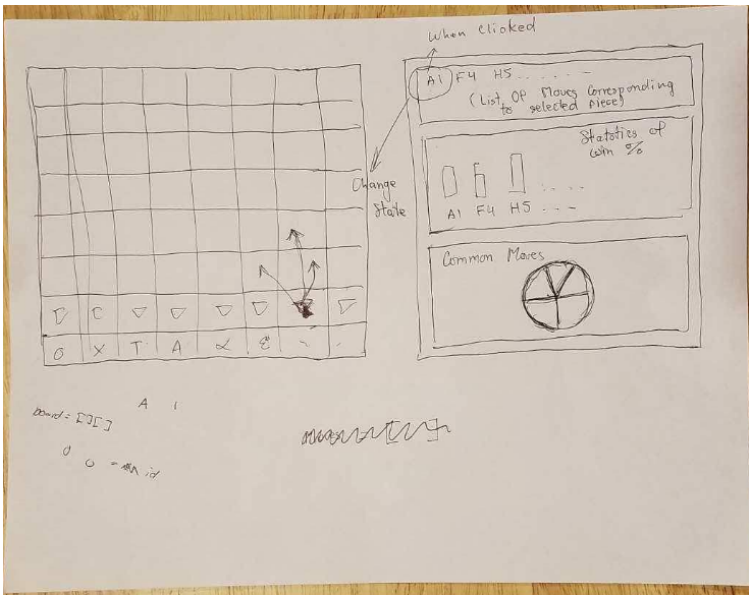
Exploratory Data Analysis What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

The data wasn't viewed with any visualizations initially. It was merely opened in a document viewer and processed with the naked eye. This data was used in order to figure out which visualizations can help answer the questions asked. Having all of the data for all of the games, it was easy to figure out that it can be parsed to find out the most common moves throughout all the different matches in the dataset. Knowing this, figuring out the percentages by win and lose states is also just as simple.

Design Evolution

The main visualizations that were chosen are a chessboard and a sidebox that would display information about potential moves. The chessboard shows the different pieces that the user can select in order to view information about it. On the right side, there is a bar chart, pie chart, and a line chart that shows information about what the next best possible move would be. These moves are calculated by parsing through an abundance of professional games and whether that move led to them winning the game.





Initially our visualizations contained only four different visuals in order to show the user the data. These four visuals are the pie chart, the bar chart, the list of potential next moves, and the chessboard indicating where the moves would go. As we brainstormed the different visuals that could be used, we decided to eventually merge a few ideas that were similar in them into the final visualization.

The reason the pie chart was chosen was because it shows the total relationship between the commonality of the common moves compared to each other very quickly. These charts are great to show discrete data as they can be easily compared to each other. Since we do not have to compare the chart as a whole to any other pie chart, it defeats the weakness of comparability.

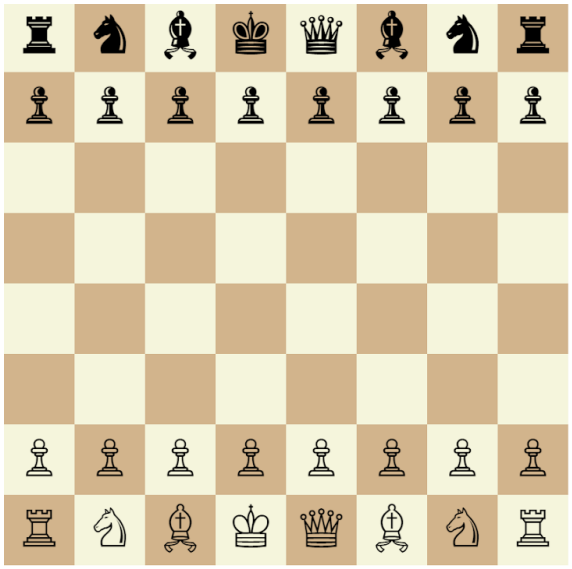
With the bar chart, since we have a finite and discrete amount of categories to show (tie, black wins, white wins), it was easiest to show it with this sort of visualization as comparability will be most simple. This was later changed to be a horizontal stacked bar chart, to make it easier to see and compare the proportions.

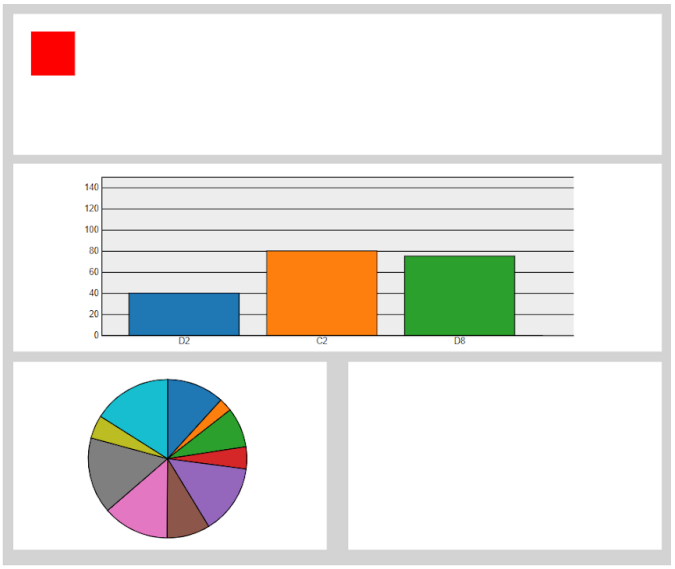
We added the text in order for users to be able to read the “raw” data for those who would rather read than see a graphical visualization. These texts would be at the top of the visualization.

Finally we have the chessboard. At first, the pieces were able to be moved around freely but was quickly changed so that the user can only move it the way it’s supposed to be moved.

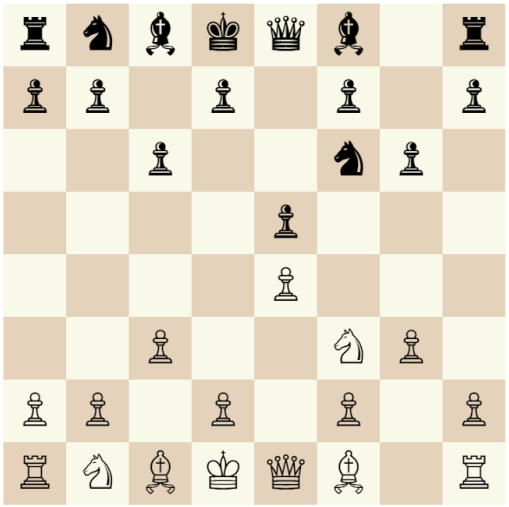
Giving the user too much freedom over moving pieces can cause potential issues with parsing through the data to find matches, especially when our data is configured as a tree.

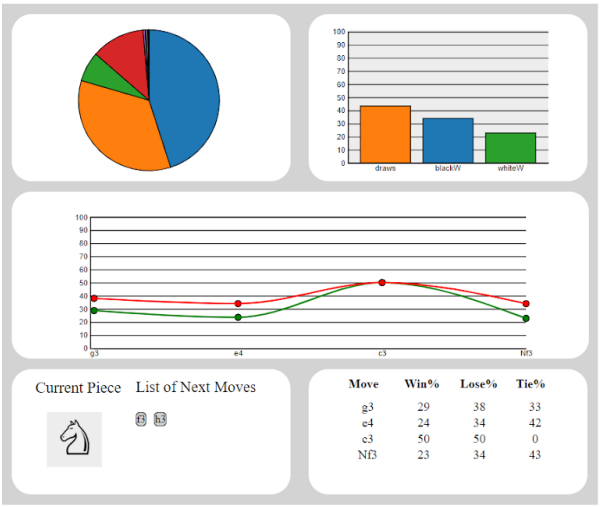
With all of this in mind, the following draft was created:





At first, there was only supposed to be the three visualizations as mentioned above as well as the chessboard. However, while working on the visualizations, it was deemed that a history of the game could prove helpful to users when visualizing the impact of their moves over time. Deviating from the proposal, textual history of the line chart, and the line chart itself is added.

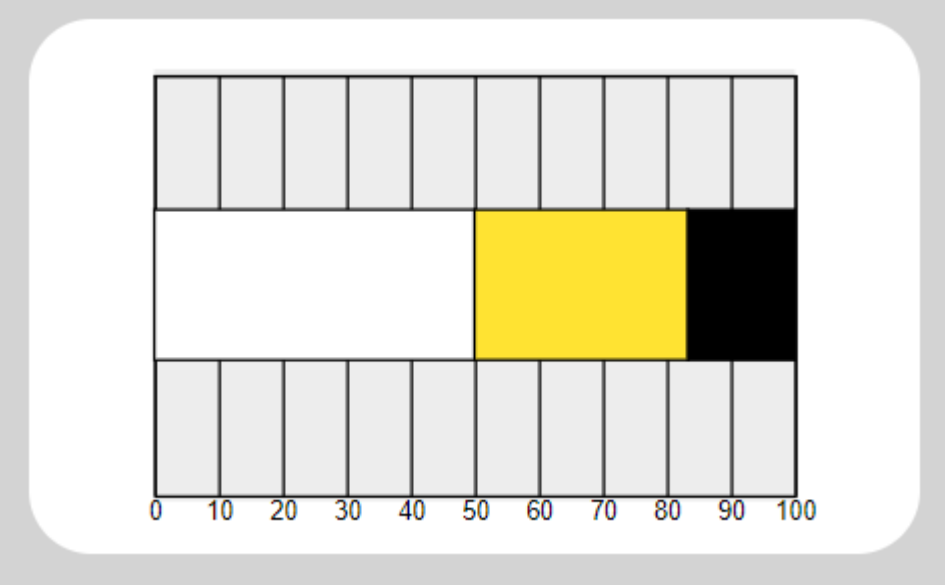




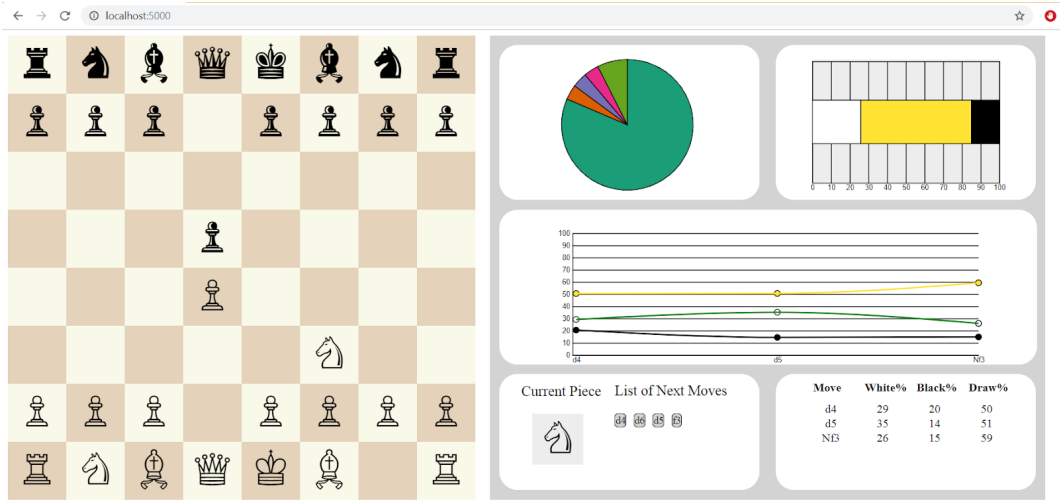
The line chart is a fantastic visualization tool to use for displaying data over time. Letting the users know how their choices affect their chances of winning over time is useful for the user to better improve their decision making or know whether to give up or continue.

The bottom right is the textual history of the line chart. With this visualization, the user can look through the history of the game and which moves were chosen.

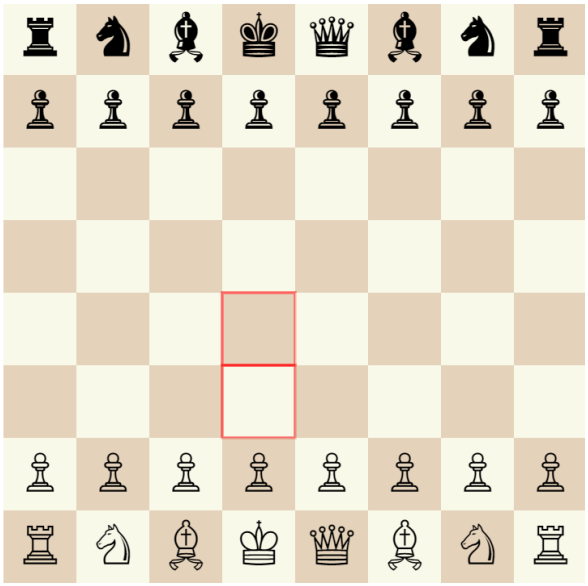
The bottom left is now where the potential moves are located. We chose not to highlight the piece on the chessboard and instead highlight it here instead to ensure that the user knows what piece is selected when pressing their next available move.



Our final change was updating the coloring of the views to make them correlate to the colors of the pieces. This helps the user more easily look at the views.



Implementation



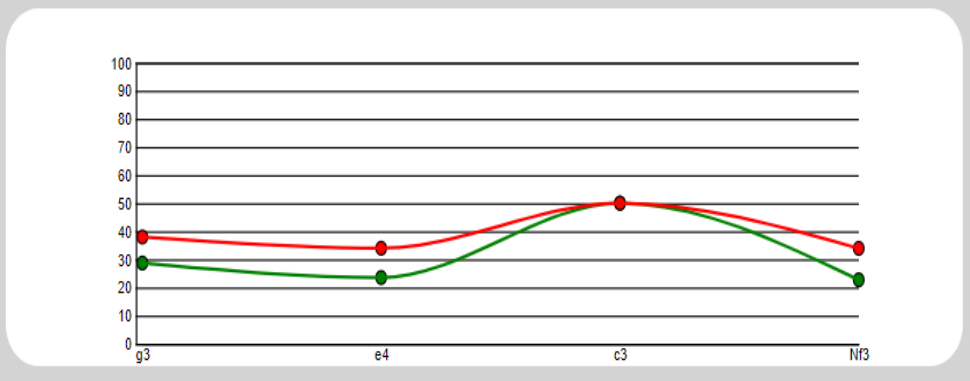
The chessboard was implemented in order for users to feel as though they're actually playing chess to mimic the actual experience as close as possible. The chessboard would show what the possible moves are by highlighting the possible squares as red, as well as show it in the box for textual purposes. As such, the chessboard will only allow you to move pieces that are matching specific states.

The pie chart's purpose is to show the user what percent a specific move is compared to the other common next moves. It is easier this way to show the users which piece they may want to analyze first. Once this piece is selected, the bar chart will update showing the win percentage of that specific move. It shows the white win rate, the black win rate, and the tie rate if that move would to be chosen.

The bar chart was chosen to represent, as a whole, the percentages on who usually wins when a specific move is made. Making the three different variables into one bar out of 100 made it easier to compare what direction the game would heavily swing in if the move is chosen.

The line plot in the center box shows the history of the win rate percentage based on the moves chosen by the user. The intent is to allow the user to track their cumulative win percentage chance over the course of all of their actions. Using this, the player can adjust their moves and learn new strategies based on the top recorded chess players around the world.

Under that, on the bottom left side, you will see a textual representation about what the next common moves are alongside the piece that you have selected, to ensure you



are reading the right data. The bottom right will show what the bar chart shows textually as well.

Evaluation What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

The visualizations helped answer some of the questions that were asked. For example, it was discovered what moves were generally detrimental to starting the game by finding out the percentage of losing when a move is considered. The most common moves were also discovered by parsing through the match data and seeing which move is the most common for the first few rounds. From then, it became increasingly harder to find matching states and data as the permutations grow in number.