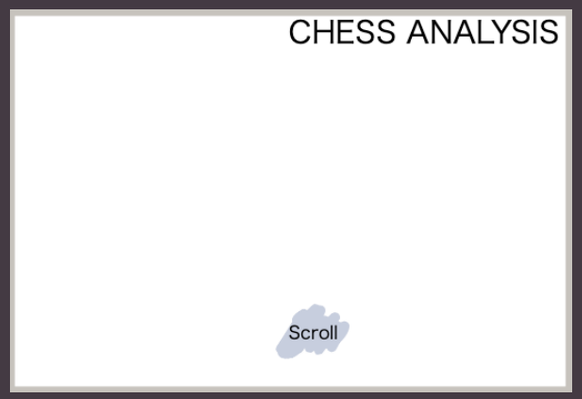
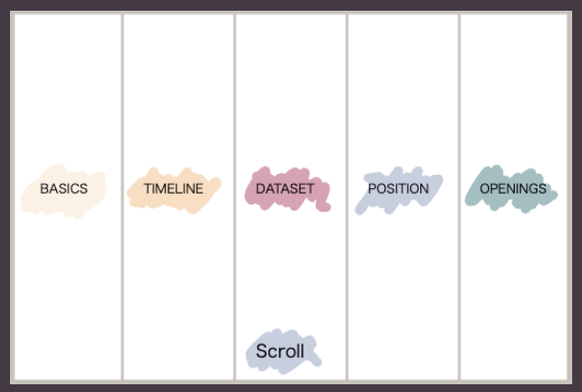
Benhaim Julien
Michelet Elisa
Vignoud Julien

THE CHESS ANALYSIS

For beginners and advanced players

INTRODUCTION





The dataset used throughout the project is a collection of **chess games** that occurred on the online chess platform called Lichess. Accounting around **one million** games, we took advantage of the dataset size to draw some tips and conclusions with a **data-driven approach**. We set our main objective towards providing an **overview** of the main aspects and strategies of chess **for each player level**. In other words, we aspire to give a general and visual understanding of the game that is tailored to users' familiarity with chess. For instance, beginners are more interested in basic tactics and common moves, intermediates with the most useful openings and advanced players with some examples of games from the best players in the dataset. The visualisations allow each player to explore and get **insights** from thousands of games, letting them take a new look at chess, use it as a tool to improve their future strategies or **spark a new interest** for the game. We will describe each visualisation individually in the next sections along with how they relate to our problematic and objective as well as the challenges that we overcame in the project development.

The **multi-level approach** is best illustrated by three visualisations: the basics, the position and the openings that are respectively best leveraged by beginner, intermediate and advanced players. While they are tailored for a particular chess background, we put an emphasis on making these visualisations **meaningful** and useful for any player.

To give an overview of the different aspects of the game we created two particular visualisations that are independent of one's chess level: a geographical timeline and graph dataset. The first is a geographical timeline that illustrates important events and interesting anecdotes in a visual way, highlighting the country in which each event occurred. By showing the **historical aspect** of chess we hope to create an interest and maybe let the user discover a new viewpoint. The second visualisation is an **alternative representation** of the dataset: we create a graph where each node is a user and each edge is a game played by two users of the dataset. The graph has been designed as a sandbox in which users can explore the games and pay attention to what they like for instance by going through games of the best players in the dataset.

The **website architecture** is another expression of our desire to adapt to any background. We didn't want to create a linear website in which every user would scroll through the visualisation one by one because some are more interesting than others depending on the player. Therefore we created a hub from which users can choose to directly go to their visualisations of interest.

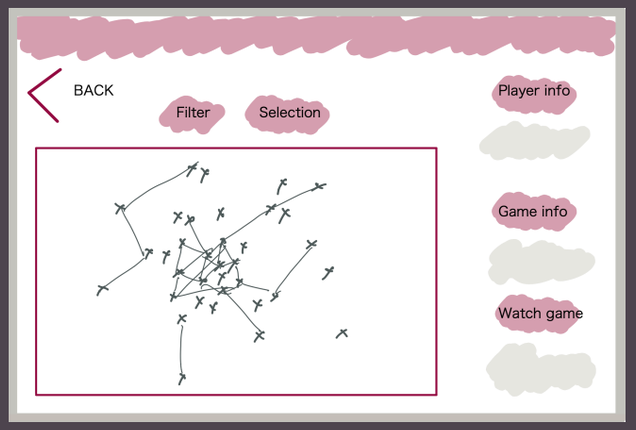
GEOGRAPHIC TIMELINE

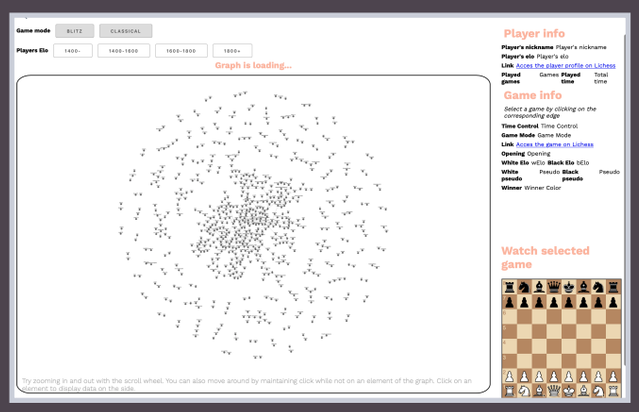
The main objective behind the idea of a geographic timeline is to create an **interest** for chess. While we cannot assume that users have the same background in chess, we cannot either assume that they are interested in the topic. Thus we decided to create a visualisation, which was completely **independent** of the chess level, focused on showing the depth of this game through its **history**. We handpicked events and anecdotes in chess history in order to combine fundamental knowledge and interesting details.

The visualisation is composed of two separate parts interacting through the events selected by the user. When they hover over an event in the timeline, the description is displayed below and the adequate country is highlighted in the world map. The user can drag and zoom the timeline to look with more clarity into their period of interest. Through this timeline, we can see the **expansion of chess** through the countries as well as the **succession of countries** where chess was most prevalent at the time.

The original idea was to create a **two-way interaction**, in which users could hover over the events and thus highlight a country, and where users could as well hover over a country to highlight the corresponding events. However, after the implementation, we realised that the country hover was causing the timeline zoom to jerk abruptly when a user was hovering quickly over multiple countries. We therefore chose to create a **colour code** for the countries that would be used in the timeline and displayed by hovering over a country, letting one easily notice events from each country. The biggest challenge of this visualisation was the interaction between the two parts, as there are sometimes some conflicts between the different event listeners. We used the D3 geoprojection library to create the map.

DATASET EXPLORATION





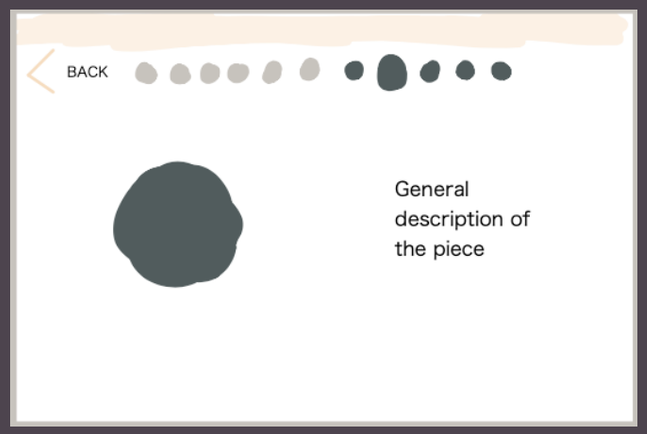
The aim of the graph visualisation is to enable the user to explore the dataset. It also gives **insights** on the structure of the Lichess platform games. We can easily see **clusters**, players who train a lot, which game modes are preferred and differences between ELO ranges. The nodes correspond to players, and the edges to chess games. When clicking on a node or an edge, the corresponding data are displayed on the side. One can see the game duration, players Elo rank, who won the game, etc. Once a game is selected, an interactive chess board can display it, move per move.

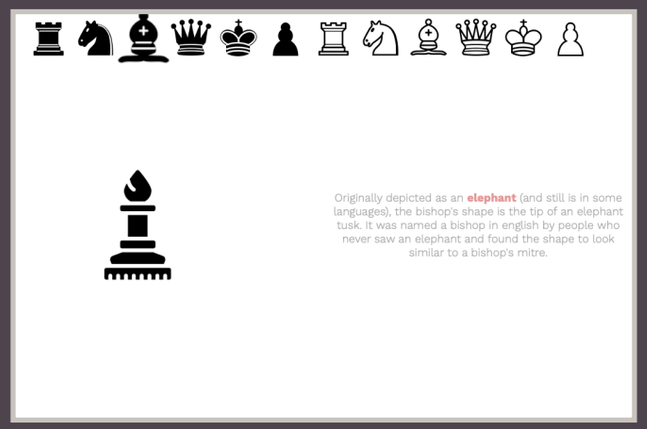
The graph is also an **interactive** way to watch thousands of different chess games, take the time to analyse them and improve. By selecting the appropriate filters, the user can watch games from an ELO range they can understand. They can also see how much time the players had to play this game as this is an important aspect of Chess.

The main changes that occurred since milestone 2 is the **degree of freedom** the user has when selecting filters. Because displaying too many games has a huge impact on performances, we had to restrict the time range and ELO range that can be displayed. Otherwise the user experience would have decreased a lot and seeing too many edges and nodes isn't meaningful anyway. Another alternative would have been to create clusters between the games depending on the zoom level, but this is much more involved and requires a lot of processing of the data. This was the main challenge faced when designing this simulation, as well as the high degree of interactivity between each element of the web page.

Beginner

BASICS



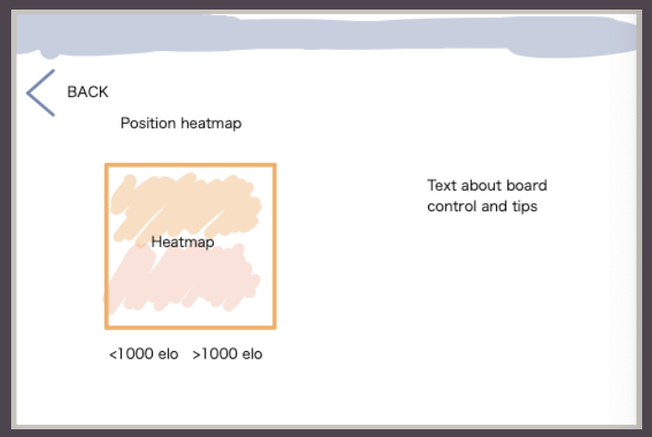


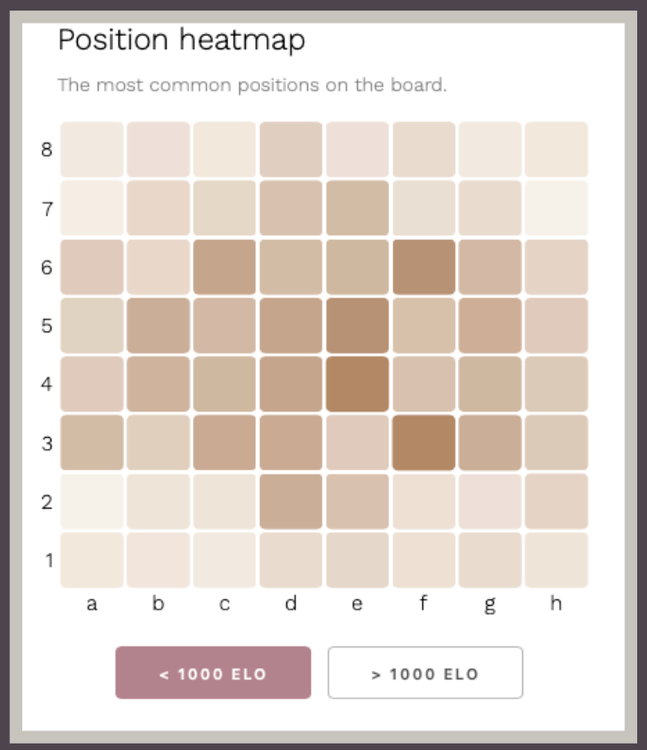
The page that describes the pieces aims to give a **fun introduction** to the rules of chess. These visualisations are targeted to a beginner audience, even though more advanced players could find some interesting facts. One can select the piece they want information on with the top menu. Then, by scrolling down the users get more and more precise facts about the piece. It starts from a general description of it, then follows descriptions of starting positions and legal moves, and finally statistics, drawn from the dataset, of the piece are presented. The whole page claims to be **interactive** by illustrating all the chessboard positions with actual pieces on a chessboard that one can play around with and soak up the game. It is useful to give an **intuition** for chess. The layout of the page also allows for a **comparative approach** that might be of interest for advanced players, comparing precise statistics on different pieces.

One challenge in creating this page laid in the computation of the statistics, finding them back from a formal notation of every game. A second difficult aspect was to find a way to display these numbers in an engaging way that allowed for interesting interpretations. The animated bar-plot is designed as an answer to that challenge, being **clear** and **easy to read**.

Intermediate

BOARD POSITIONING





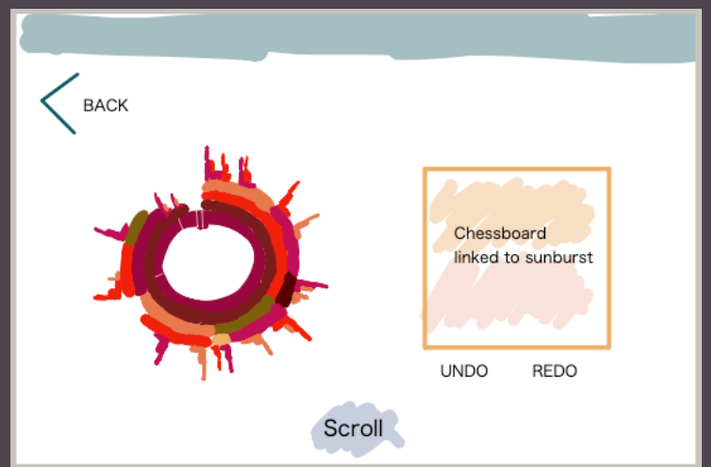
The position heat map is visualisation aimed at chess players with a minimal experience that didn't delve into the theory. With a **data-driven approach**, we draw conclusions that are backed by chess theory. That is, we show that **center control** is a main stake in the game by comparing the **game styles** of low level and high level players present in the database. In fact, we can see that higher ELO players tend to occupy the center squares more than beginners.

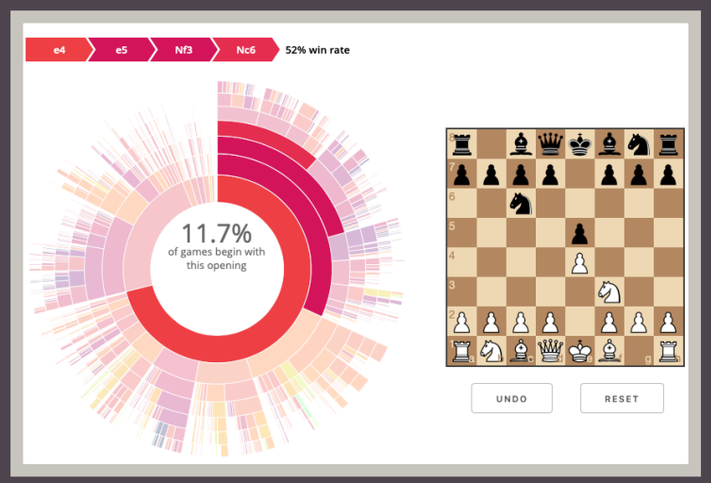
The visualisation is a heat map that displays each square of the board with a different colour and intensity, depending on how often it is occupied in the games of the database. A dark colour shows that a square is often occupied while a light colour means that it is rarely occupied. We let the user filter the games used to compute the **occupancy frequency** by the player level. By comparing the different heat maps according to player levels, the user sees an **empirical confirmation** of the theory.

The challenges of the visualisation was both in the data processing that is done beforehand and the creating a visualisation that emphasises best the difference between the two heat maps. The current display is the best **compromise of simplicity and effectiveness** by leveraging the power of D3 tooltip to display additional information only on command.

Advanced

OPENINGS DIVERSITY





The opening sunburst is the first and main plot that has been in our mind. Openings is a fascinating domain of chess and being able to let the users explore the **concrete habits** of all the players in the database is exciting. The more advanced players will make the most use of this visualisation but a lot of effort has been made to make it meaningful to any background. Thanks to its **versatility**, users can both explore the opening tree and try their own sequence of moves. Moreover we give some insights that are backed by theory and illustrated by the visualisation.

The visualisation is composed of two parts that interact both ways. The left part is a **radial layout** of the most common sequences of move. The i-th ring is a partition of the i-th possible moves, where each arc has an area proportional to the frequency of the move. An arc colour denotes the win rate of this particular sequence of move in the database. Hovering over an arc displays more information such as the opening name or the precise win percentage. The right part is a **chessboard** that illustrates the pieces arrangement when a user hovers over an opening. One can also manually change the piece disposition on the board to find more information on a particular opening.

This visualisation presented a lot of challenges due to both the complexity of the sunburst plot and the **tight interactions** between the two parts. We made use of the D3 hierarchy library for the radial layout. The chessboard was originally designed to be static and only display the selected opening. One could not play a particular opening on the chessboard. The only way to find a particular opening required to know the Portable Game Notation and explore the tree. That was in contradiction with our goal to make the visualisation meaningful to anyone therefore we decided to add another layer of complexity for us by making the chessboard interactive and thus providing a **smoother experience** for users.

Benhaim Julien

My main contribution to this project is the exploration of the dataset through the graph. It started with some data pre-processing work on Python to turn the list of games into an easily readable JSON file describing a graph. Then I started to work on cytoscape.js according to what was advised in the graph lecture. I tried to bring a lot of interactivity to make this viz interesting for the user. The main challenge was performances.

Other than that, I explored the possibility to use Vue.JS on this website, to have better code structure and better handling of events and data updates. But the overhead of learning to use it for the team wasn't worth the clarity we would have gained. I also tried to create an interactive list of the games, that the user could filter and search in to explore the dataset, which could have been integrated within the graph or other visualisation. But even with libraries made for handling huge data lists such as clusterize.js , the performance and interactivity level were not satisfying. Vue.js might also have helped with this problem.

Michelet Elisa

I first did the exploratory data analysis to understand the dataset in depth. On the website, I took care of the page for the beginners. I started by computing on the dataset the statistics for each piece and plotted them as an animated bar plot with d3 to get myself familiar with the library, new to me. Still for this page, I researched good ways to describe each piece, textually (the descriptions) and visually (the configuration of chessboards). I also drew all the sketches used throughout the project and created the final video.

Vignoud Julien

The main contributions I brought to the project are the openings and the timeline visualisations, which required a significant load of work. Through those, I learnt to use the D3 library in-depth, particularly the hierarchy and geo-projection modules, as well as design complex interactions between different plots. I also created the position heat map, which was the first visualisation I created and was a good way of getting to know more about the dataset. Finally I designed the website, in which I demonstrate the use of CSS animation.