

Étude des ouvertures aux échecs

DataViz - Document de cadrage

Maxime ANTOINE, Salim CHABCHOUB, Mustapha EL ABRIDI,
Jessica NDIAYE, Dorian TONNIS



Problématique

Notre projet vise à mieux comprendre le rôle de l'ouverture dans le résultat des parties d'échecs en ligne, ainsi que son impact sur la précision des joueurs.

Plus précisément, nous souhaitons répondre aux questions suivantes :

- Quelles sont les ouvertures les plus jouées et quels résultats donnent-elles ?
- Quelle est la précision des joueurs lors de l'ouverture ?
- En perdant l'avantage dans l'ouverture, à quelle fréquence puis-je revenir dans la partie ? Est-il encore possible de gagner ?

À qui s'adresse la visualisation, quelles tâches seront effectuées au travers de votre projet ?

La visualisation s'adresse à des joueurs d'échecs désireux de progresser dans leurs ouvertures, de construire un répertoire, et d'avoir une meilleure compréhension du spectre d'ouvertures possibles.

Quelques questions auxquelles les diagrammes permettent de répondre :

- Quelles sont les ouvertures les plus jouées ? Pour quel résultat ? Quelles sont les défenses les plus jouées contre mon ouverture ?
- Quelles sont les ouvertures les plus adaptées à mon niveau et à la cadence de la partie ?
- À quel point l'ouverture est-elle décisive dans la partie ?
- Quelle est la précision des joueurs lors de l'ouverture ?

Répondre à ces questions est important pour la progression du joueur, car cela permet de construire un répertoire d'ouvertures cohérent avec son niveau et durable dans le temps.

Avoir un bon répertoire peut permettre de prendre un avantage immédiat dès le début de la partie, ce qui peut favoriser la progression du joueur.

Sources de données choisies

Pour ce projet, nous nous baserons sur les données open source de Lichess :

<https://database.lichess.org/>



Une des problématiques liées à ces données est leur quantité. Au total, plus de 2 To de données sont disponibles, et chacun des derniers mois de données pèse environ 30 Go (par exemple, octobre 2025 : 29,9 Go).

Le traitement de ces données nécessite beaucoup de ressources et de stockage dont nous ne disposons pas vraiment.

C'est pourquoi, dans un premier temps, nous allons nous restreindre à l'analyse des données de 2013 (environ 500 Mo), puis nous envisagerons potentiellement d'étendre notre analyse à plus de données si le temps nous le permet.

Voici à quoi ressemble les données d'une partie sur Lichess :

```

[Event "Rated Bullet tournament https://lichess.org/tournament/yc1WW2Ox"]
[Site "https://lichess.org/PpwPOZMq"]
[Date "2017.04.01"]
[Round "-"]
[White "Abbot"]
[Black "Costello"]
[Result "0-1"]
[UTCDate "2017.04.01"]
[UTCTime "11:32:01"]
[WhiteElo "2100"]
[BlackElo "2000"]
[WhiteRatingDiff "-4"]
[BlackRatingDiff "+1"]
[WhiteTitle "FM"]
[ECO "B30"]
[Opening "Sicilian Defense: Old Sicilian"]
[TimeControl "300+0"]
[Termination "Time forfeit"]

1. e4 { [%eval 0.17] [%clk 0:00:30] } 1... c5 { [%eval 0.19] [%clk 0:00:30] }
2. Nf3 { [%eval 0.25] [%clk 0:00:29] } 2... Nc6 { [%eval 0.33] [%clk 0:00:30] }
3. Bc4 { [%eval -0.13] [%clk 0:00:28] } 3... e6 { [%eval -0.04] [%clk 0:00:30] }
4. c3 { [%eval -0.4] [%clk 0:00:27] } 4... b5? { [%eval 1.18] [%clk 0:00:30] }
5. Bb3?! { [%eval 0.21] [%clk 0:00:26] } 5... c4 { [%eval 0.32] [%clk 0:00:29] }
6. Bc2 { [%eval 0.2] [%clk 0:00:25] } 6... a5 { [%eval 0.6] [%clk 0:00:29] }
7. d4 { [%eval 0.29] [%clk 0:00:23] } 7... cxd3 { [%eval 0.6] [%clk 0:00:27] }
8. Qxd3 { [%eval 0.12] [%clk 0:00:22] } 8... Nf6 { [%eval 0.52] [%clk 0:00:26] }
9. e5 { [%eval 0.39] [%clk 0:00:21] } 9... Nd5 { [%eval 0.45] [%clk 0:00:25] }
10. Bg5?! { [%eval -0.44] [%clk 0:00:18] } 10... Qc7 { [%eval -0.12] [%clk 0:00:23] }
11. Nbd2?? { [%eval -3.15] [%clk 0:00:14] } 11... h6 { [%eval -2.99] [%clk 0:00:23] }
12. Bh4 { [%eval -3.0] [%clk 0:00:11] } 12... Ba6? { [%eval -0.12] [%clk 0:00:23] }
13. b3?? { [%eval -4.14] [%clk 0:00:02] } 13... Nf4? { [%eval -2.73] [%clk 0:00:21] } 0-1

```

Parmi ces informations, nous allons avoir besoin des champs suivants :

- UTCDate : date de la partie
- Result : résultat de la partie (1-0, 0-1, ½-½)
- whiteElo : classement des Blancs
- blackElo : classement des Noirs
- ECO : référence de l'ouverture dans : Encyclopedia of Chess Openings
- Opening : Nom de l'ouverture
- TimeControl: Cadence (rapide, blitz, bullet)
- moves : liste des coups

Une autre difficulté liée aux données est l'évaluation de l'ordinateur. Nous avons besoin de cette information afin de déterminer la précision des joueurs, mais elle n'est présente que sur environ 6 % des parties selon Lichess (puisque cela nécessite une analyse par le moteur). Il faudra donc filtrer les données pour isoler les parties qui contiennent les évaluations dont nous avons besoin.

En cas de problème lié à l'exploitation des données Lichess, nous avons un plan B :

<https://database.nikonoel.fr/>

Cette base de données est basée sur les données de Lichess mais pré-traitée. L'auteur a extrait uniquement les parties de joueurs de haut niveau et a exclu les parties en bullet. Cela pourrait être une option de secours. L'inconvénient de ce jeu de données est que l'on ne

pourra pas faire de comparatif de niveau puisqu'il n'y a que des parties de haut niveau.

Travaux importants liés au projet

En plus de fournir les données des parties jouées, Lichess montre quelques exemples de projets réalisés avec celles-ci. Nous allons en présenter quelques-uns.

[ChessRoots](#)

ChessRoots est un projet très similaire à notre visualisation sur les ouvertures les plus jouées. Il consiste en une application web montrant un arbre avec tous les coups jouables à partir de la situation actuelle. On peut construire l'arbre de la partie en sélectionnant à chaque tour un des coups proposés.

Les points forts de ce projet sont que l'on peut visualiser le déroulé de la partie grâce à l'arbre et au plateau se mettant automatiquement à jour en fonction des coups joués. On peut également savoir le nombre de parties présentes dans la base de données avec cette configuration spécifique.

Cependant, malgré le fait que l'épaisseur de la flèche soit proportionnelle au nombre d'occurrences du coup, il est compliqué de se rendre compte en un seul coup d'œil de la fréquence associée à chaque ouverture.

[Chessbook](#)

Chessbook est une application éducative gratuite conçue pour aider les joueurs d'échecs à établir un répertoire d'ouvertures solide. Elle permet de créer un répertoire personnalisé, d'entraîner par répétition espacée et d'analyser automatiquement les lacunes dans le jeu d'un joueur. Chessbook utilise la technique de la répétition espacée , qui est scientifiquement prouvée pour mémoriser les ouvertures rapidement . L'application lui permet de combiner des ouvertures issues de différentes sources afin de bâtir un répertoire vraiment unique.

Son principal atout est son approche pédagogique : elle ne se limite pas à montrer les ouvertures, mais accompagne le joueur dans leur assimilation et leur pratique régulière.

Il évalue aussi les positions du répertoire dans l'objectif de couverture du joueur puis en prenant une moyenne pondérée de ces positions, le poids étant la fréquence attendue de ces positions à votre niveau.

En revanche, Chessbook reste centrée sur l'entraînement individuel et ne fournit pas de visualisation statistique globale des ouvertures (par fréquence, efficacité selon l'Elo ou la cadence). Cela limite la compréhension du contexte général des ouvertures, que notre projet cherche justement à mettre en lumière.

[Chess Opening Repertoire Builder](#)

Cette application pensée pour aider à développer, comprendre en profondeur et pratiquer les répertoires d'ouvertures aux échecs. Elle nous permet d'explorer visuellement différentes lignes d'ouverture, d'annoter les coups directement sur le graphique, et d'effectuer des recherches dans la base de données officielle des ouvertures. On peut également tester nos connaissances grâce à des quiz dédiés et jouer des parties à partir de positions prédefinies pour nous entraîner de manière ciblée.

Cette application, à l'instar du projet précédent, conserve une dimension pédagogique, mais sous un angle plus ludique. Elle propose divers graphiques illustrant les ouvertures.

Utilisation des ouvertures en fonction de l'elo

L'utilisation des ouvertures en fonction de l'elo est essentielle pour les joueurs d'échecs , car elle permet de se conformer à leur style de jeu et aux compétences qu'ils ont . Par exemple , les joueurs qui préfèrent un contrôle solide sur l'échiquier dès les premiers coups optent pour les ouvertures classiques comme l'ouverture espagnole .

Selon les travaux de *Brendan-Budok Durand-Le Ludec* , "Diversifier son répertoire, c'est aussi le faire évoluer. Si le répertoire doit s'adapter au niveau du joueur, il est normal qu'il évolue au fil de sa progression." Son premier constat est que les performances de la grande majorité des variantes ne sont pas stables en fonction des niveaux . Cela confirme que les amateurs devrait se concentrer sur jouer des ouvertures qui leur donne des points en fonction du niveau des adversaires qu'ils rencontrent.

Pour être plus spécifique :

- Les catégories <2000 Elo : utilisent plus les gambits et variantes , qui sont particulièrement efficaces à ce niveau. Elles offrent de bonnes chances de gain grâce à l'effet de surprise et à la difficulté des joueurs moins expérimentés à défendre correctement . Ainsi, les systèmes simples sont adaptés , car ils sont faciles à maîtriser et apportent une efficacité immédiate.
- Les catégories ≥ 2000 Elo : Les joueurs expérimentés savent neutraliser les surprises et exploitent les imprécisions adverses. Les ouvertures complexes, notamment avec les Noirs, deviennent alors un meilleur choix, car elles permettent de jouer véritablement pour le gain en s'appuyant sur une préparation solide.

Choix et construction du répertoire d'ouvertures – Mémoire DEFFE

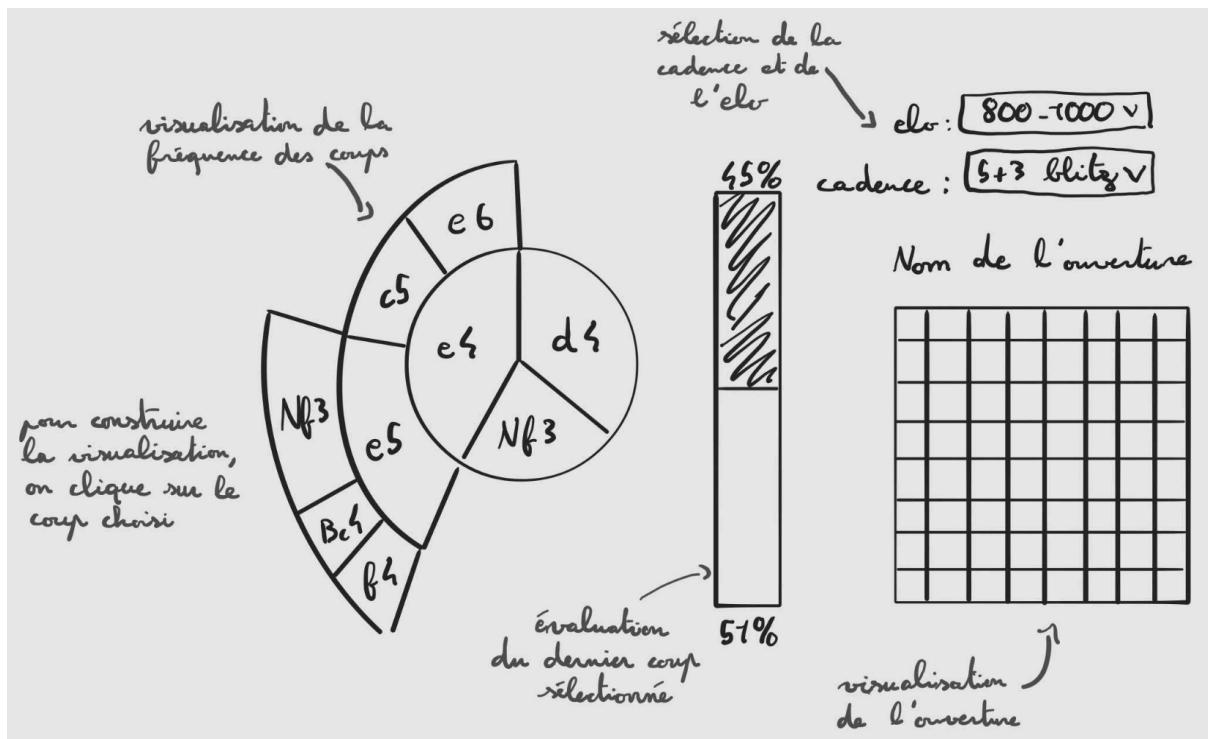
Organisation

Pour la communication de l'équipe, nous avons opté pour Discord. Concernant le partage de code, nous utiliserons GitHub, cette plateforme que nous maîtrisons déjà grâce à nos expériences passées, ce qui facilitera l'échange de notre travail.

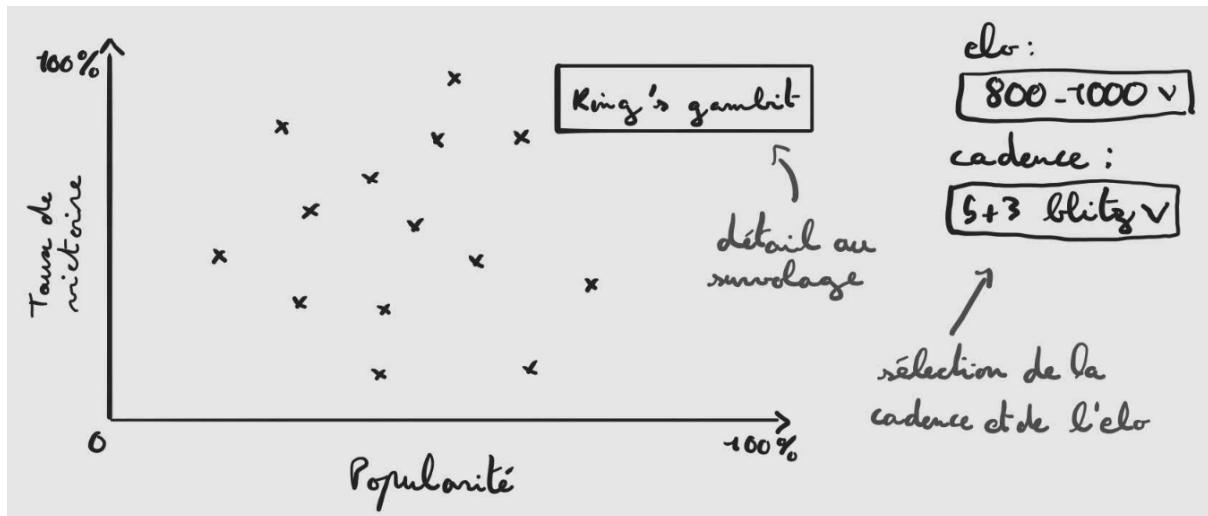
Nous travaillons actuellement la plupart du temps à distance. Nous avons cependant organisé quelques sessions de travail collaboratif :

- **Première session (26/11) :** Consacrée à la génération d'idées de graphiques. Nous avons utilisé la technique du 'pliage de feuille en huit' pour que chacun propose huit graphiques différents.

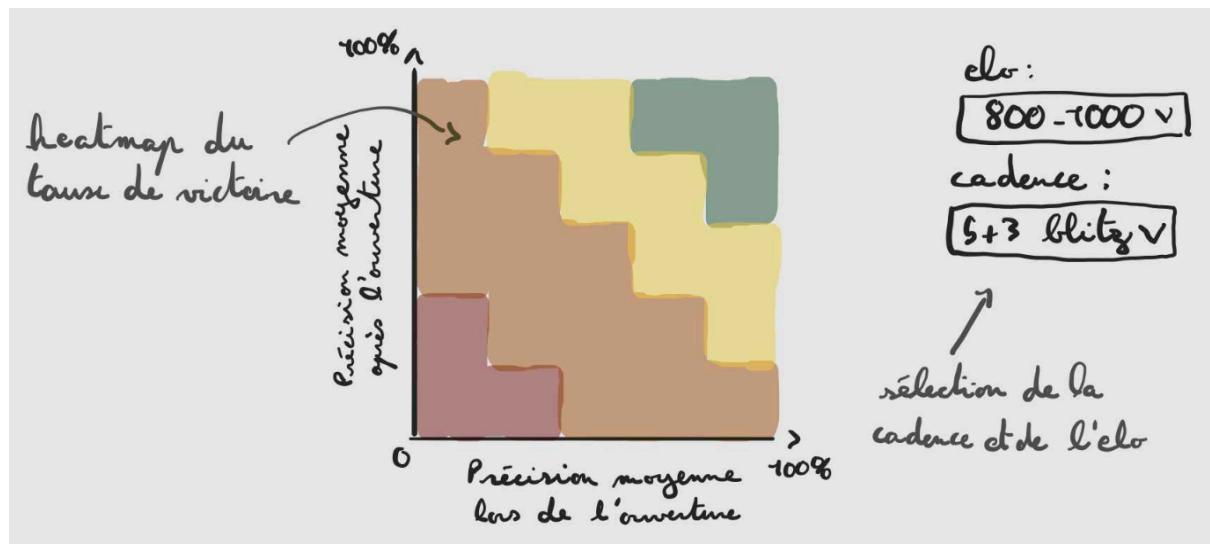
Scan des esquisses finales



Visualisation n°1 : Visualisation des ouvertures les plus jouées et de leur efficacité par elo et par cadence.



Visualisation n°2 : Visualisation de la popularité et du taux de victoire des ouvertures.



Visualisation n°3 : Visualisation du taux de victoire en fonction de la précision lors de l'ouverture et au cours du reste de la partie.