



Etude et analyse des parties d'échecs

Sommaire

I. Introduction sur les échecs	p.3
II. Base de données.....	p.3
III. Analyse des parties.....	p.4
a. Influence de l'écart du niveau des joueurs sur leur parties.....	p.4
b. Niveau des joueurs	p.6
c. Les différents types d'ouverture	p.6
d. Nombre de coups et fréquence des ouvertures	p.8
e. Relations entre les ouvertures et le niveau du joueur	p.10
f. Type de parties jouées	p.12
g. Issue de la partie.....	p.12
IV. Conclusion	p.13
V. Bibliographie.....	p.13

I. Introduction sur les échecs

Les échecs sont l'un des plus anciens jeux de société qui se joue à deux. En s'exportant dans le monde depuis l'Asie, celui-ci a évolué en Europe afin de prendre la forme du jeu actuel que l'on connaît. On peut y jouer entre amis, en compétition, en physique ou même en ligne désormais grâce aux technologies actuelles. Il est composé d'un échiquier de 64 cases numérotés de a1 à h8 et de 32 pions noirs et 32 pions blancs.

Pour y jouer, chaque joueur se voit attribuer une couleur et un temps est défini en début de partie, par exemple pas de limites, 10min de jeu maximum par joueur, ou alors même en plus un joueur peut récupérer 5 secondes par exemple à la fin de son tour.

Par la suite, le joueur blanc commence en faisant une ouverture (parmi des milliers existante). L'objectif est d'essayer d'affaiblir le plus possible son adversaire en mangeant ses pions sans mettre son roi en danger.

Pour que la partie se finisse, il faut que le roi soit mis en échec et mat, que le temps imparti pour un joueur soit écoulé, ou qu'il y ait match nul (pat). Cela survient lorsqu'un joueur ne peut plus déplacer aucun pion sans mettre son roi en échec.

Dans l'objectif de mieux comprendre comment sont jouées les parties, nous avons récupéré et analysé une banque de données afin d'en tirer quels impacts ont les différents facteurs sur l'issue de la partie et la manière de jouer.

De nos jours les échecs ont beaucoup évolué avec le temps. En effet, avec l'arrivée de grosse plateforme en ligne d'échec tel que Chess.com (qui fut l'une de nos sources principales pour cette étude) le nombre de joueur s'élève à 600 000 000. De plus, pour départager tous ces joueurs un système de élo a été mise en place. Ce système permet de déterminer si le niveau d'un joueur est plus ou moins élevé et pouvoir dans le même temps accéder à des compétitions plus ou moins importantes. En moyenne, pour chaque partie gagnée, le joueur victorieux augmente de 7 élo et celui de son adversaire baisse de 7 élo. Ce chiffre est basé sur un match fait entre deux joueurs à niveaux équivalent. Evidemment, si un joueur qui a un écart de 1000 élo de plus et que ce même joueur perd il va dans le même perdre plus d'élo. Aujourd'hui Garry Kasparov est l'un des joueurs avec le plus d'elo, son score s'élève à plus de 2800.

II. Base de données

Afin de réaliser notre étude sur ce jeu, nous avons récupéré sur le site Kaggle une banque de données basée sur 20058 parties effectuées sur le site chess.com. De ce fichier nous avons pu en tirer diverses informations :

- Résultat du joueur : victoire, défaite, abandon ou nul qui seront représenté dans notre tableau par TRUE, FALSE ou DRAW.
- Couleur du joueur : blanc ou noire et pour un souci de facilité d'études nous avons convertis ces valeurs en binaire (1 pour blanc et 0 pour noire)
- Nombre de coups dans une partie du joueur

- Durée de la partie : Chaque partie possède 3 modes de jeu (pour des sessions en ligne). Des parties Bullet = 1min, blitz = 3min, Rapid = 10min et les Daily qui peuvent aller jusqu'à 4jours.
- Le type d'ouverture réalisée : Il existe plusieurs types d'ouverture "classique" et à partir desquelles sont issus différentes variantes. Dans cette étude nous nous concentrerons plus dans les ouvertures dites classique.
- Le nombre de coups pour effectuer l'ouverture : Le nombre de coup sont des valeurs quantitatives
- Le niveau du joueur : Le niveau d'un joueur est en élo et est une valeur quantitative

Ces données, qui aux premiers abords peuvent sembler indépendantes les unes des autres, nous permettent en réalité de pouvoir obtenir une mine d'information sur la manière de jouer des joueurs en fonction de plusieurs paramètres que nous allons voir.

Afin d'en tirer au mieux partie, nous allons ainsi comparer ces données et les mettre en corrélation en comparant les résultats en fonction du niveau et de la couleur attribuée, le nombre de coups en fonction de l'écart de niveau. De plus, nous avons rajouté différentes colonnes dans notre base de données afin de pouvoir étudié au mieux notre data set. Tout d'abord nous avons calculer la différence de niveau entre les joueurs pour chaque partie. Puis, à partir de cette colonne nous avons créé une colonne "niveau justifié" que l'on détaillera plus en détail dans notre mémoire.

III. Analyse des parties

a. Influence de l'écart du niveau des joueurs sur leur parties

Dans cette partie nous allons voir en quoi le niveau des joueurs influe sur leur qualité de jeu et les différents choix fait durant leurs parties. Pour étudier cela, nous avons regardé chaque partie, et voir si le niveau du joueur était "justifié". Tout d'abord qu'est-ce qu'un niveau justifié ? On considère un niveau justifié pour une partie si la personne qui a gagné cette partie avait un niveau supérieur à l'autre personne.

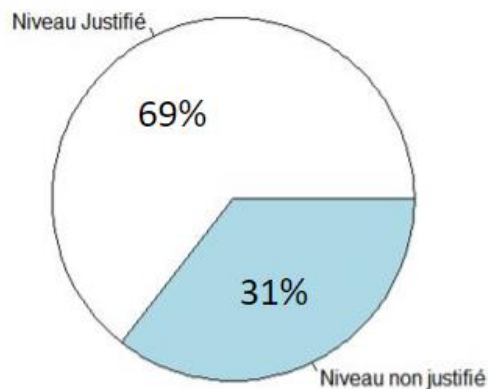
Cependant, dans notre base de données les informations qui nous était donnée était : la couleur gagnante (Blanc, Noir ou nul) et le niveau de chaque couleur (Exemple : Classement noir = 1200 Elo).

Nous avons donc dans un premier temps converti ces valeurs qualitatives en valeur quantitative pour pouvoir utiliser les différents diagrammes. Nous avons ensuite regardé la différence de niveau entre chaque joueur pour chaque partie.

Par la suite, nous avons attribué à chaque partie une valeur qualitative qui nous disait si le niveau était justifié ou pas.

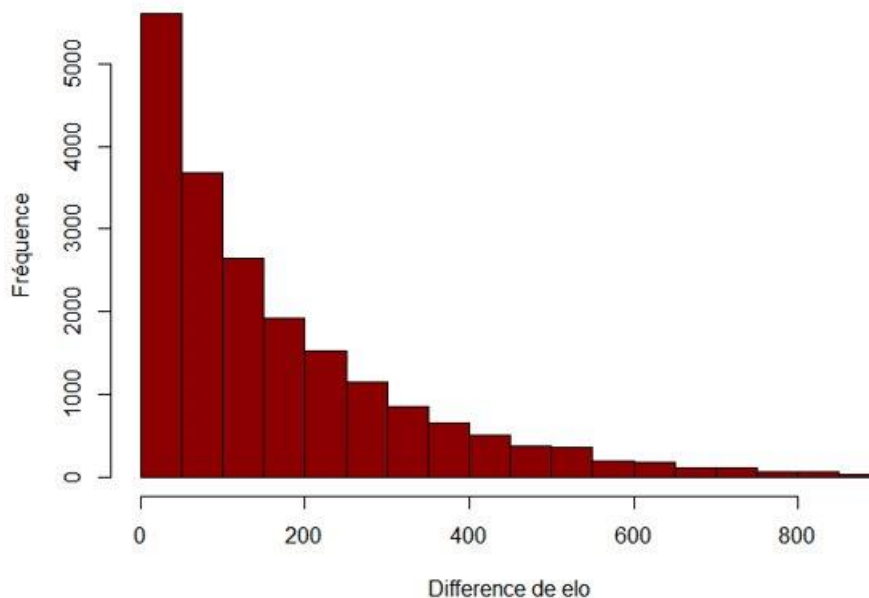
Nous avons répertorié toutes ces données dans un camembert suivant :

La différence de niveau est elle justifié



Ce que l'on remarque dans ce camembert c'est que le niveau du joueur contribue en partie à la réussite ou non de ce joueur. Cependant, cette différence de niveau ne fait pas tout. En effet 31% des parties ont été gagnées par le joueur ayant un niveau inférieur à son adversaire. On peut donc en déduire que la différence de niveau ne donne pas directement le joueur gagnant. On remarque aussi que ces 69% de n

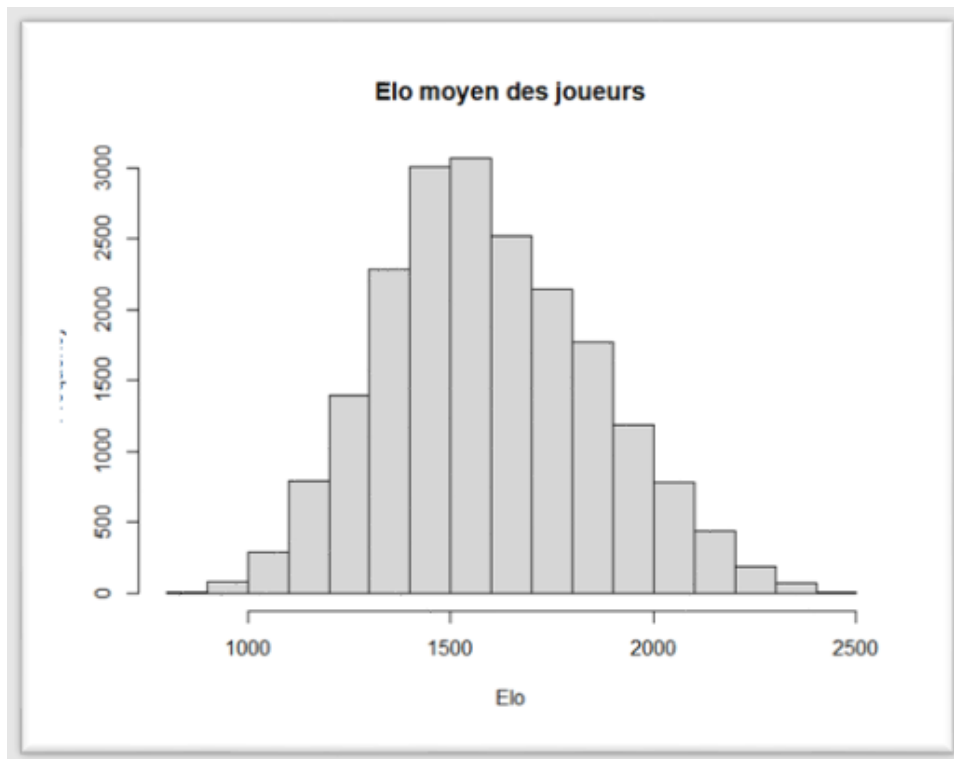
Représentation de la différence de niveau



On en a déduit que la différence de Elo entre chaque joueur était en général assez faible (proche de 0). Ce qui fait que les 39% de niveau non justifié peut être en partie biaisé par la faible différence de niveau des joueurs. En effet, si on regarde la moyenne de la différence de niveau entre les joueurs sur chaque partie on trouve 19,76 Elo soit proche de 0.

b. Niveau des joueurs

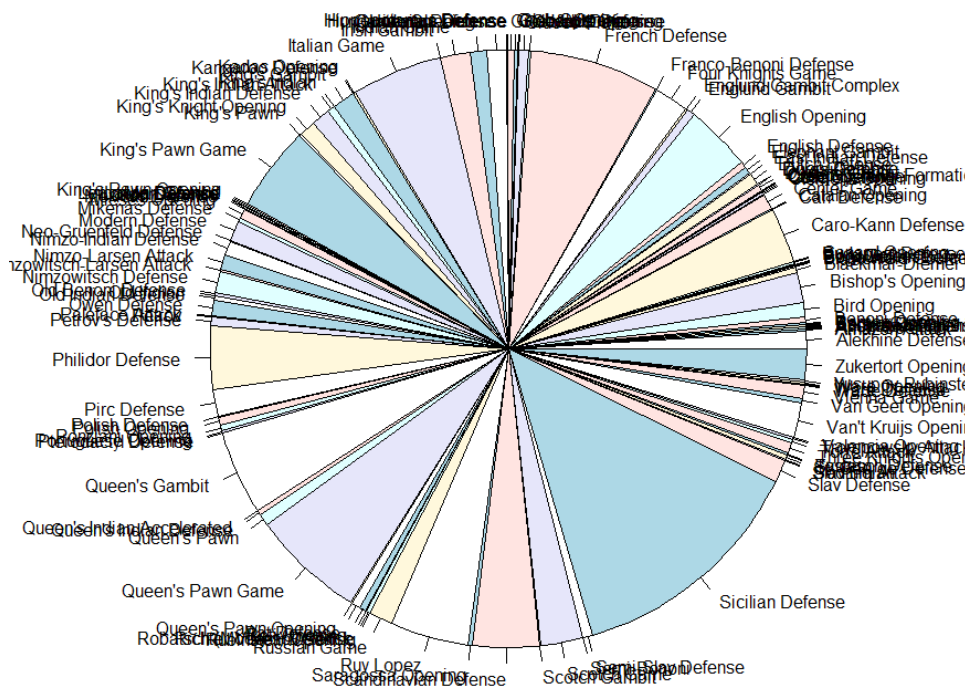
Nous nous sommes donc penchés sur les différentes techniques utilisées par les joueurs expérimentés et nous avons relevé plusieurs points. Tout d'abord quel était le niveau moyen des joueurs ? On a répertorié ces différents niveaux sur l'histogramme ci-dessous.



On en déduit que la répartition des différents niveaux sur l'ensemble des parties suit une loi normale centrée en 1500 Elo. On a donc calculé le niveau médian des joueurs avec la commande "médiane(niveau)" et on a trouvé que le niveau médian des joueurs était de 1567, avec une variance de 84 697. On a donc une grande concentration des différents niveaux autour de 1500, ce que l'on retrouve notamment dans l'histogramme de différence de niveau.

c. Les différents types d'ouverture

Intéressons-nous ensuite aux nombres de mouvement pour une ouverture. En général, une partie d'échec dépend grandement de son ouverture car elle va ensuite déduire tout le déroulement de la partie. Aux échecs il existe des centaines de types différents d'ouverture avec pour chaque type d'ouverture de multiples variantes. Un joueur d'échec expérimenté se sera renseigné par ces différentes variantes et connaît une grande majorité des ouvertures. Cependant ce n'est pas le cas de tout le monde comme nous le montre ce camembert.



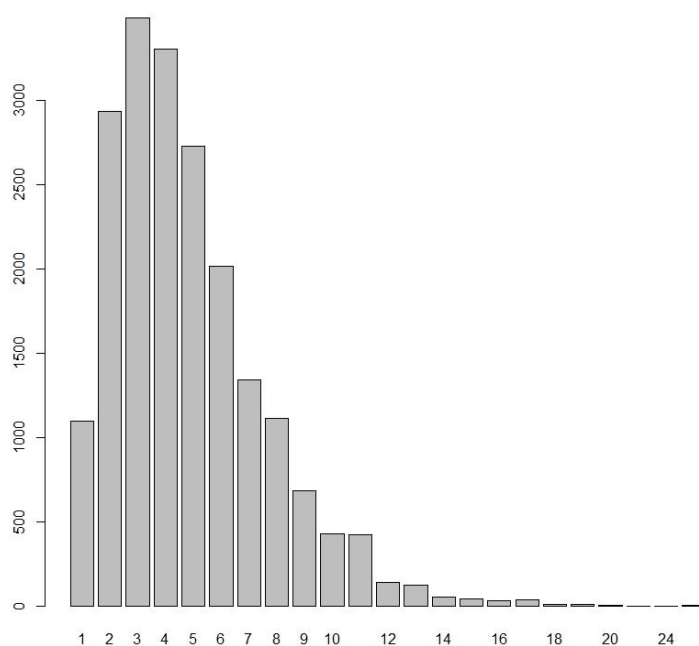
Ce camembert peu dans un premier temps paraître complexe, cependant il permet de mettre en avant la fréquence des différents types d'ouverture et d'en extraire les principales. On remarque quelque chose d'assez intrigant, c'est que la défense sicilienne (qui est la première ouverture que l'on apprend aux échecs) est l'ouverture la plus jouée suivi de la French Défense qui est pareil une ouverture dites "classique". On peut donc en déduire qu'une grande partie des joueurs utilise la défense sicilienne lors de leur partie. Cette ouverture correspond bien aux diagrammes que l'on a vu juste avant car il pour un élo médian de 1500 (ce qui correspond à un niveau amateur), l'ouverture choisi sera elle aussi de la même difficulté. En effet, la défense sicilienne n'est uniquement composée que de 2 coups ce qui en fait une ouverture assez simple.

A l'inverse, les ouvertures les moins jouées sont celles nécessitant le moins de coup et qui sont jouées par des joueurs ayant un élo plutôt élevé.

On peut en conclure que pour progresser dans les échecs il faut apprendre plusieurs coups, permettant d'avoir une meilleure attaque contre son adversaire sans mettre en danger son roi.

On retrouve cet aspect lorsque l'on étudie notamment le nombre de coup à l'ouverture comme dans ce bar plot.

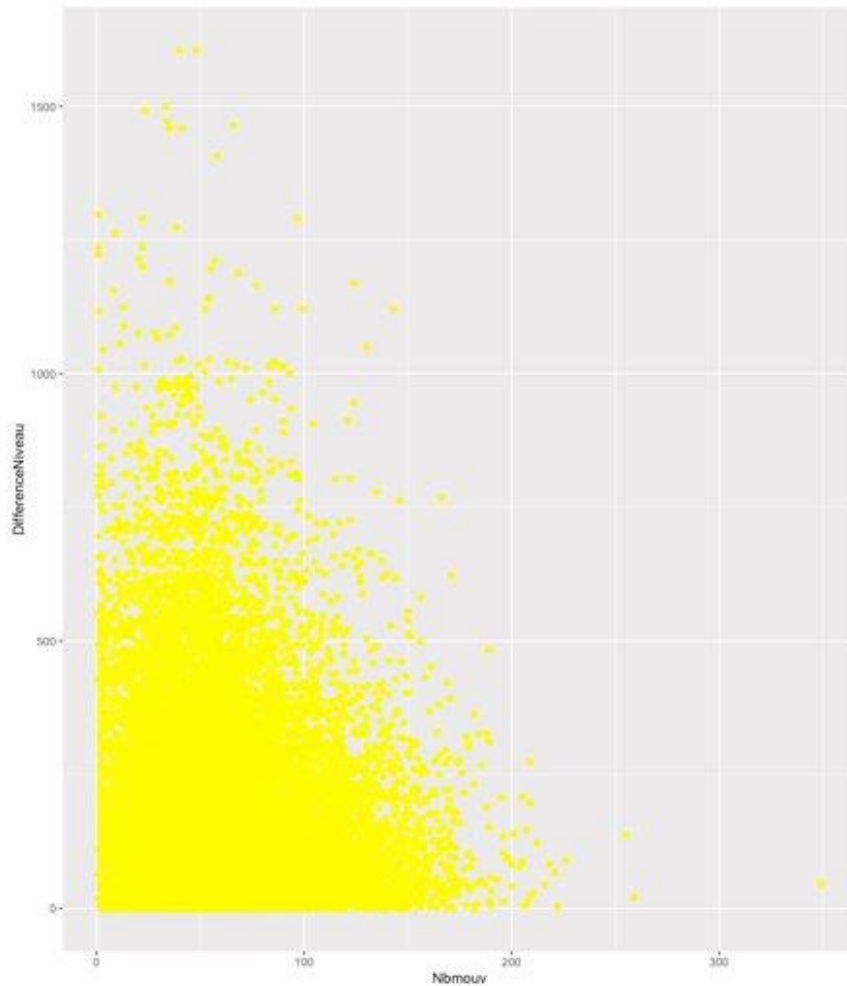
d. Nombre de coups et fréquence des ouvertures



On remarque notamment que le nombre de coup pour une ouverture tourne autour des 4 ou 5 coups en moyenne ce qui correspond à des ouvertures dites classique comme la défense sicilienne.

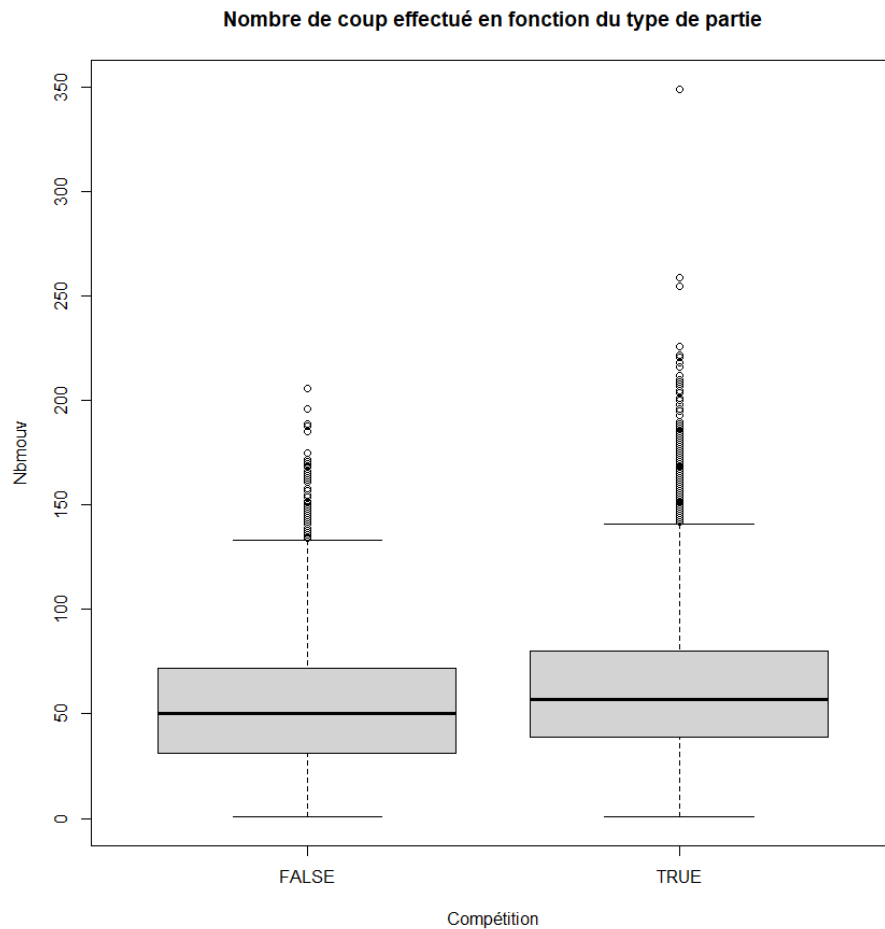
Cependant, cette corrélation est aussi liée aux nombres de coup pour une ouverture. En effet, comme on peut le voir dans ce nuage de point, les ouvertures qu'elles soit classique ou non ne dépasse rarement la ligne des 10 coups.

e. Relations entre les ouvertures et le niveau du joueur



On y remarque deux points assez essentiels. Tout d'abord, on a un amas de points concentrés sur une différence de niveau entre 0 et 500 et un nombre de mouvement entre 0 et 100. Cela est en partie lié à ce que l'on a vu sur les diagrammes précédents qui est que la différence de niveau entre les joueurs pour chaque partie est assez faible et que le niveau moyen est lui aussi amateur. Cependant on remarque aussi qu'un nombre de mouvement élevé est liée à une différence de niveau élevé et réciproquement un nombre de mouvement proche de 0 est lié à une différence de niveau élevé. On peut en conclure que lorsque les matchs sont à niveaux équivalents la partie est beaucoup plus longues et technique. On remarque aussi que ce nuage de point est justifié à travers un rapport de corrélation qui est négatif ($\text{cor}(\text{Différence de niveau}, \text{Nombre mouvement}) = -0,111$).

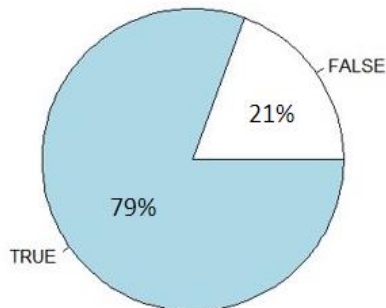
Ce nombre de coup varie aussi en fonction de la nature de la partie effectuée. Comme on le voit dans la boîte à moustaches si dessous.



On remarque que le nombre de coup pour une partie est plus important en compétition (TRUE) qu'en match amicale (FALSE). On peut donc en déduire que les joueurs sont moins concentrés et porte moins d'attention à des parties hors compétition qu'en compétition. Cette boîte à moustache permet de rebondir sur un autre point qui est lui aussi assez important qui est la proportion de partie en compétition comme nous l'indique ce camembert si dessous.

f. Type de parties jouées

Proportion de partie en compétition



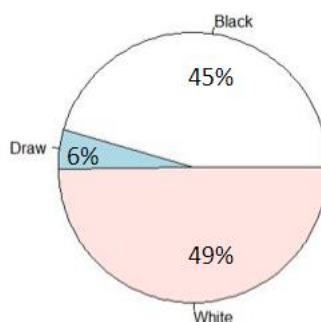
On remarque que la majorité des parties se font en compétition et dont le résultat influe sur le nombre de joueur. En effet, 79% des parties se font en compétitions ce qui est assez considérable. On peut expliquer cela par la pertinence des parties en compétitions. En effet, un joueur ayant un niveau amateur (1500 Elo) aura envie de monter en niveau, et pour cela il aura besoin de parties plus complexe où le joueur pourrait apprendre plus.

Question annexe : Est-ce que la couleur choisie par un joueur influe sur le résultat de la partie ?

Nous nous sommes posé cette question et avons mis les différents résultats dans le camembert suivant.

g. Issue de la partie

Couleur gagnante



On remarque que le fait de commencé (donc d'avoir la couleur blanche), influe légèrement sur le résultat de la partie. C'est pour cela que la couleur de chaque joueur dans chaque partie est tirée au sort car sinon les joueurs choisiraient tous la couleur blanche pour profiter de cet avantage. Cependant cela représente un écart unique de 4% sur un ensemble de parties très précis. On peut donc supposer que cet écart a tendance à se rétrécir pour un nombre de parties étudié plus important.

IV. Conclusion

En conclusion, nous avons remarqué que l'issue d'une partie d'échecs est influencée par plusieurs facteurs tel que la couleur, le niveau, l'ouverture, le nombre de coup du joueur. Ainsi ces données permettent à un joueur qui souhaite progresser de savoir quels types de coups jouer, combien de coups faire, ses chances de gagner. Cependant la quantité de donnée que nous avons n'est pas suffisante pour en faire une étude approfondie afin d'obtenir les informations sur par exemple les coups prévus à l'avance, les fins de parties, avoir des statistiques plus précises sur certains niveaux de jeux tel que les personnes avec un grand élo car ce sont des joueurs plus rares.

V. Bibliographie

Banque de donnée :

<https://www.kaggle.com/datasets/mysarahmadbhat/online-chess-games>

Histoire des échecs et manière d'y jouer :

[https://fr.wikipedia.org/wiki/%C3%89checs#Fin de la partie](https://fr.wikipedia.org/wiki/%C3%89checs#Fin_de_la_partie)

https://fr.wikipedia.org/wiki/Histoire_du_jeu_d%27%C3%A9checs

<https://www.echecs.club/strategies/>

<https://www.ichess.net/blog/chess-strategy-tips-for-beginners/>

<https://www.chess.com/fr/article/view/combien-y-a-t-il-de-joueurs-d-echecs-dans-le-monde>

[https://fr.wikipedia.org/wiki/Joueurs_d%27%C3%A9checs_ayant_obtenu_un_classement Elo de 2 700 points ou plus#:~:text=En%202021%2C%20seuls%20quatorze%20joueurs,5%20mois%20en%20novembre%202021%20\).](https://fr.wikipedia.org/wiki/Joueurs_d%27%C3%A9checs_ayant_obtenu_un_classement_Elo_de_2_700_points_ou_plus#:~:text=En%202021%2C%20seuls%20quatorze%20joueurs,5%20mois%20en%20novembre%202021%20).)