An Exploratory Analysis On Chess.com Grandmasters

Yipeng Zhao, Nathanael Nemia, Ahmed Mosaad

2022-12-15

R Markdown

INTRO

Every year chess.com releases hundreds of thousands of games played by GM's (Grandmasters) on their platform. Using this data some questions will be explored such as: what is the distribution of game types played, what rating of players play the most frequently, what methods are usually used to end games. Please keep in mind that this data is only from grandmasters who played regardless of their rating. Sometimes they may play on lower rated accounts etc. which may distort the data, so, some conclusions may not be entirely accurate.

```
## # A tibble: 6 x 12
##   time_control end_time            rated time_cl~1 rules gm_us~2 white~3 white~4
##   <chr>        <dttm>              <lgl> <chr>     <chr> <chr>   <chr>     <dbl>
## 1 300+1        2021-06-19 11:40:40 TRUE  blitz     chess 123lt   vaisha~    2658
## 2 300+1        2021-06-19 11:50:06 TRUE  blitz     chess 123lt   123lt      2627
## 3 300+1        2021-06-19 12:01:17 TRUE  blitz     chess 123lt   vaisha~    2641
## 4 300+1        2021-06-19 12:13:05 TRUE  blitz     chess 123lt   123lt      2629
## 5 300+1        2021-06-19 12:28:54 TRUE  blitz     chess 123lt   vaisha~    2657
## 6 300+1        2021-06-19 12:41:31 TRUE  blitz     chess 123lt   123lt      2631
## # ... with 4 more variables: white_result <chr>, black_username <chr>,
## #   black_rating <dbl>, black_result <chr>, and abbreviated variable names
## #   1: time_class, 2: gm_username, 3: white_username, 4: white_rating
```
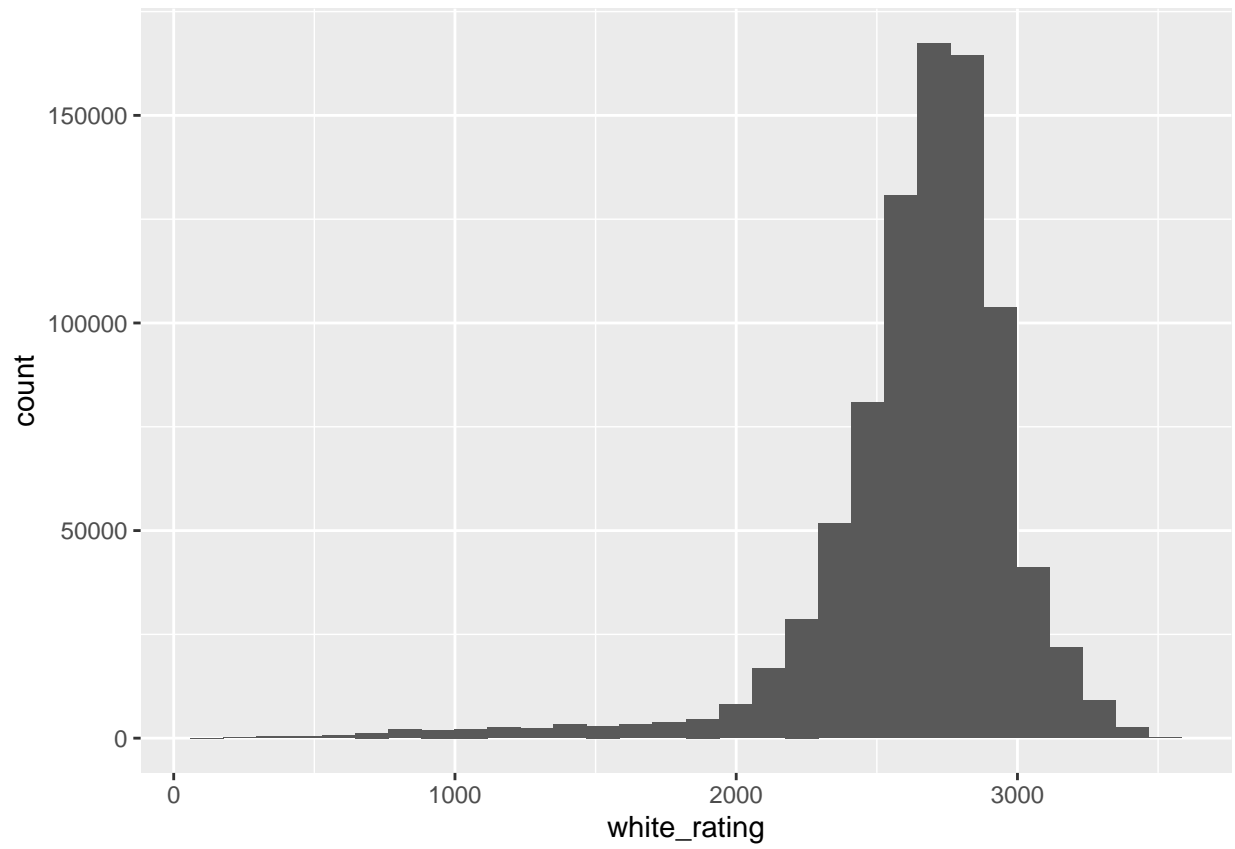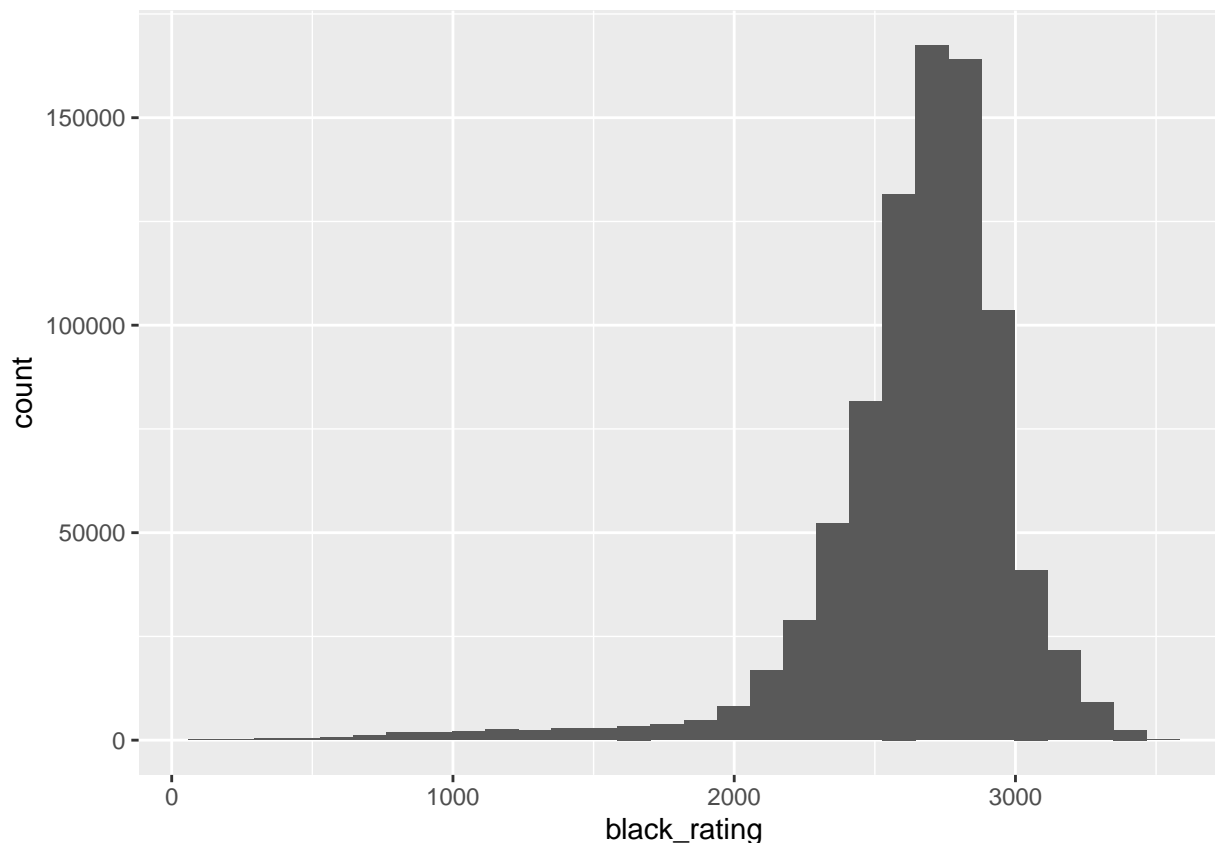
The figures above simply show games played on the chess.com website from the year 2021 by GM's. A GM is a chess player who has achieved 2500 points in their career, they need not maintain that number. Even though the average rating of users on chess.com is around 1400, the average GM is about 2500. These graphs show the distribution of games played based on rating with most games played by GM's at about 2600 rating. The symmetry of both figures shows the evenness of the chess.com matchmaking system.

```
## # A tibble: 2 x 2
##   rated      n
##   <lgl>  <int>
## 1 FALSE  45801
## 2 TRUE  813569
```

```
## # A tibble: 10 x 2
##    white_result          n
##    <chr>             <int>
##  1 win              409809
##  2 resigned         201052
##  3 timeout          101316
##  4 checkmated        60620
##  5 repetition        31532
##  6 agreed            16073
##  7 insufficient      13227
##  8 timevsinsufficient 11863
##  9 bughousepartnerlose  6440
## 10 stalemate          4145
```

```
## # A tibble: 10 x 2
```

```
##    black_result           n
##    <chr>              <int>
##  1 win               372454
##  2 resigned          230204
##  3 timeout           102777
##  4 checkmated         68216
##  5 repetition         31532
##  6 agreed             16073
##  7 insufficient       13227
##  8 timevsinsufficient 11863
##  9 bughousepartnerlose 5814
## 10 stalemate           4145
```

This table shows whether the game played was rated as well as the distribution of ways the game ended. You can see that the majority of games played are rated.
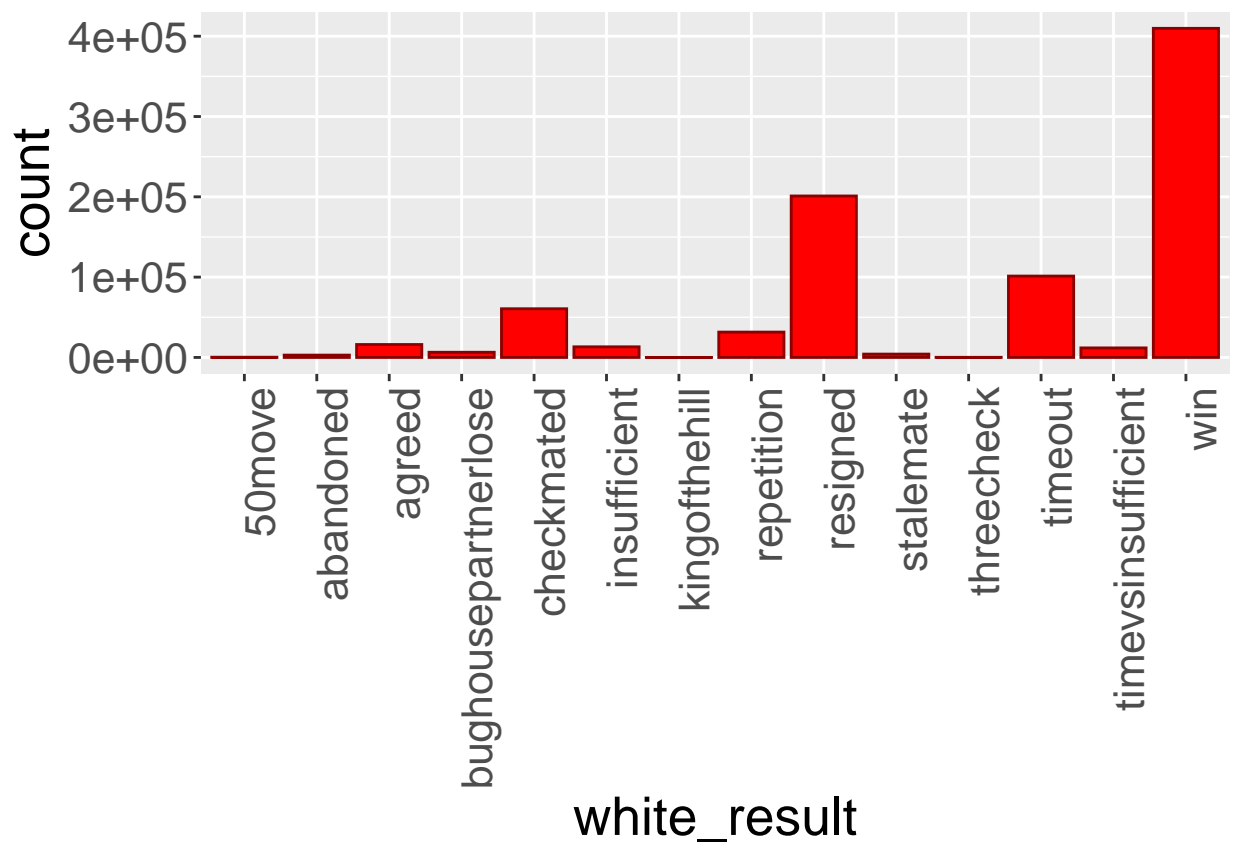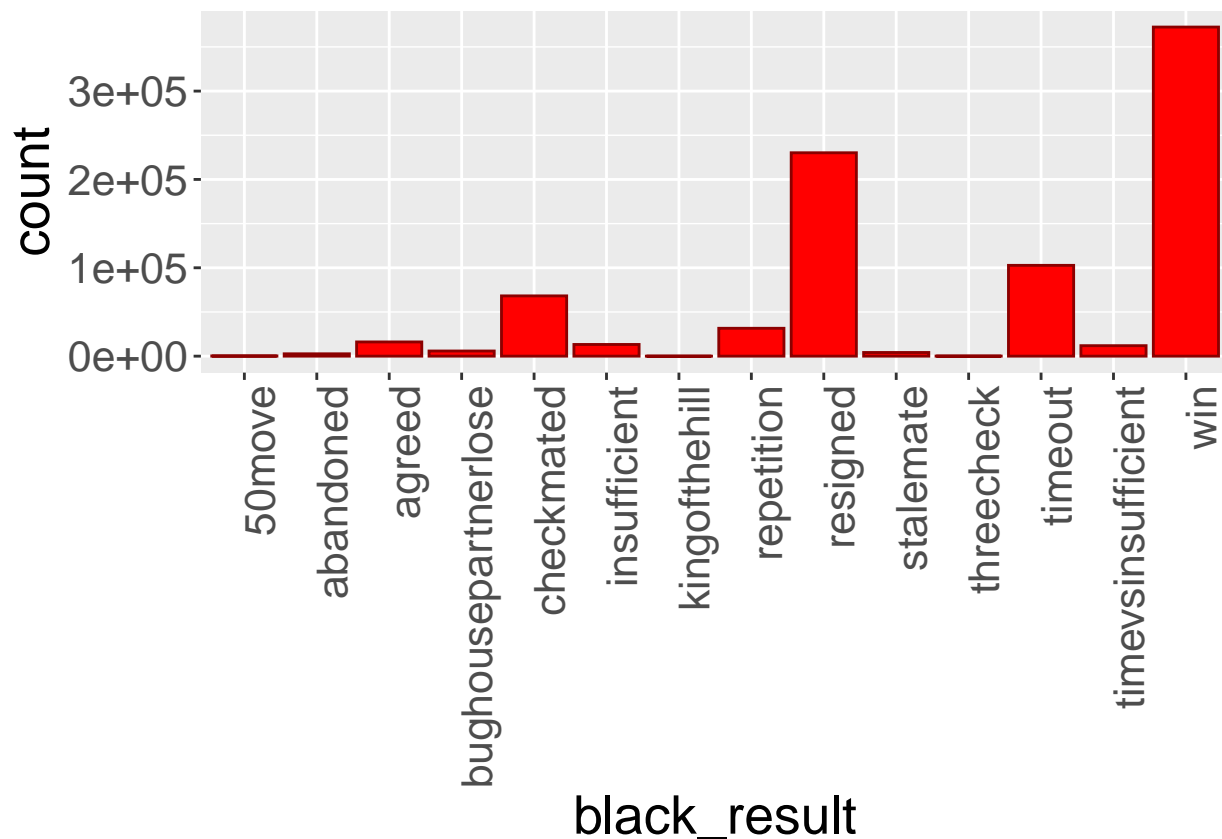
The data above gives insight into the slight advantage of playing with the white pieces. White's win rate is at 50% while black is at 47%, the difference is due to a fundamental rule where the white pieces always make the first move. Also, it should be noted that a considerable amount of games end in a timeout as well as resignation which shows the understanding the GM's have to know when they have lost before getting checkmated.

```
## # A tibble: 4 x 2
##   time_class       n
##   <chr>        <int>
## 1 blitz       599394
## 2 bullet      238246
## 3 daily         3268
## 4 rapid        18462

## # A tibble: 10 x 2
##    time_control      n
##    <chr>         <int>
##  1 180          505841
##  2 60           165947
##  3 30            59834
##  4 180+1         49626
##  5 180+2         31160
##  6 300            7558
##  7 600            7443
##  8 10             4567
##  9 60+1           2992
## 10 900+10         1972
```

This table shows the distribution of games played in the given time class as well as the type of time control they use.
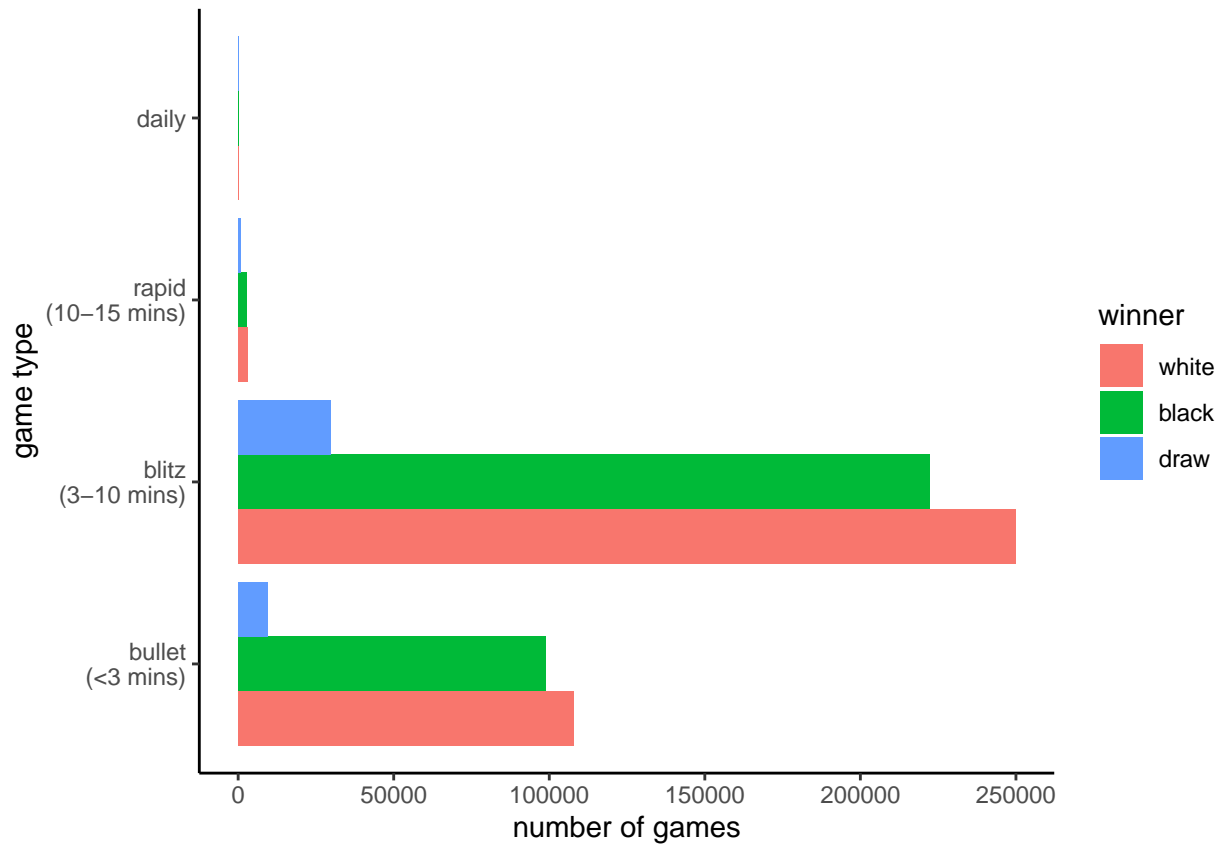
This figure focuses on the time control, which is how long each player has per game. The most common games are the blitz and bullet formats which make the games that can last for 10 minutes. It can be said that GM's favor shorter games in online chess rather than the longer draining classical time formats. Again, the white pieces have a slight advantage for making the first move.

Conclusion

Chess is a complex game but when looked at from the perspective of statistics there are some insights that simplify the sport. Through the distribution of the GM's chess games we can see chess for what it is without the bias of inexperienced players' data. The slightly higher win percentage with the white pieces and the favoring of shorter timed games lead us to believe that chess is more fun when players can rely on instinct and are pressured by time rather than thinking 13 or 14 moves ahead with games lasting hours long.