Estimating chess playing strength

DATA ANALYST NANODEGREE

JOSTEIN LANGSTRAND, PHD

2020-05-26

Project outline

This analysis aims to produce a model that is capable of predicting a chess player's strength in terms of rating, based on the information from played games.

Data from games played on one of the largest chess sites lichess.org

Games from a large online tournament, the *yearly classical arena*, that was played on May 15th 2020.

About 16,000 games were played in the tournament. Roughly 5,000 contain computer evaluation, and were used for the analysis.

Variables

The purpose of this project is to identify variables that can predict playing strength. The estimate for playing strength (dependent variable) will be the players' actual ratings. The assumption is that the following variables may contribute to predicting playing strength. They will therefore be used as predictors (independent variables):

- Average centipawn loss (aCPL)

- Rating difference between players

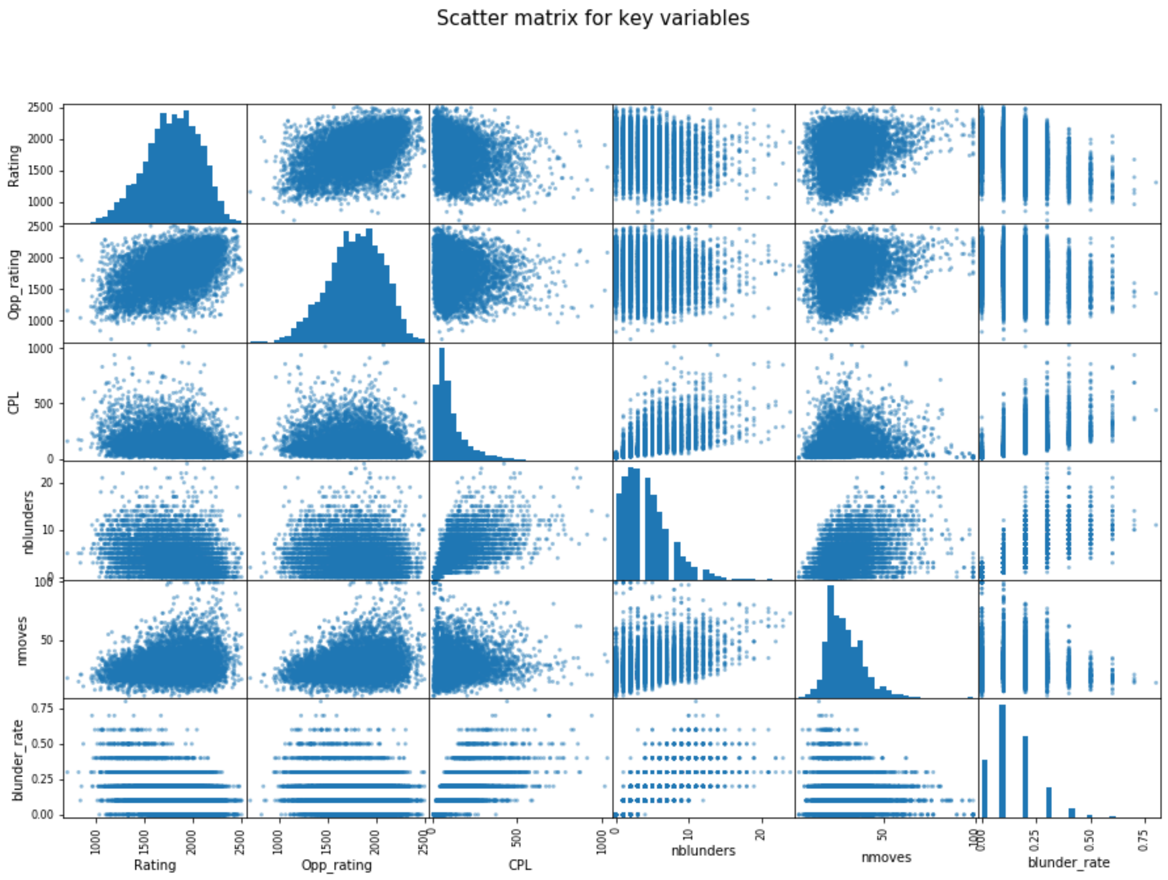
- Number of moves played

- Number of blunders (big mistakes)

- Blunder rate (number of blunders / number of moves)

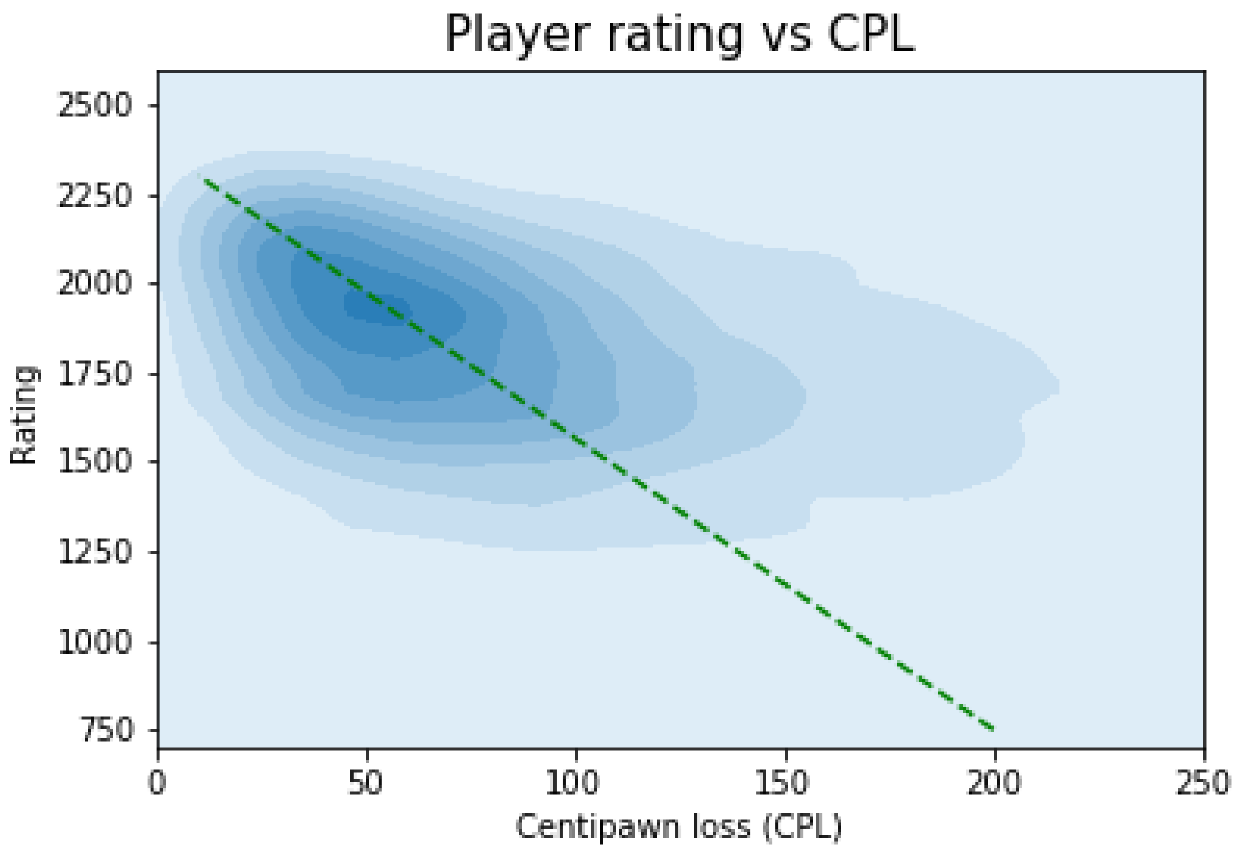
Correlations between variables

- There is a clear correlation between player ratings
- The rating distribution is fairly symmetrical
- Other variables have asymmetrical distributions
- Apart from player ratings, other correlations are quite weak.



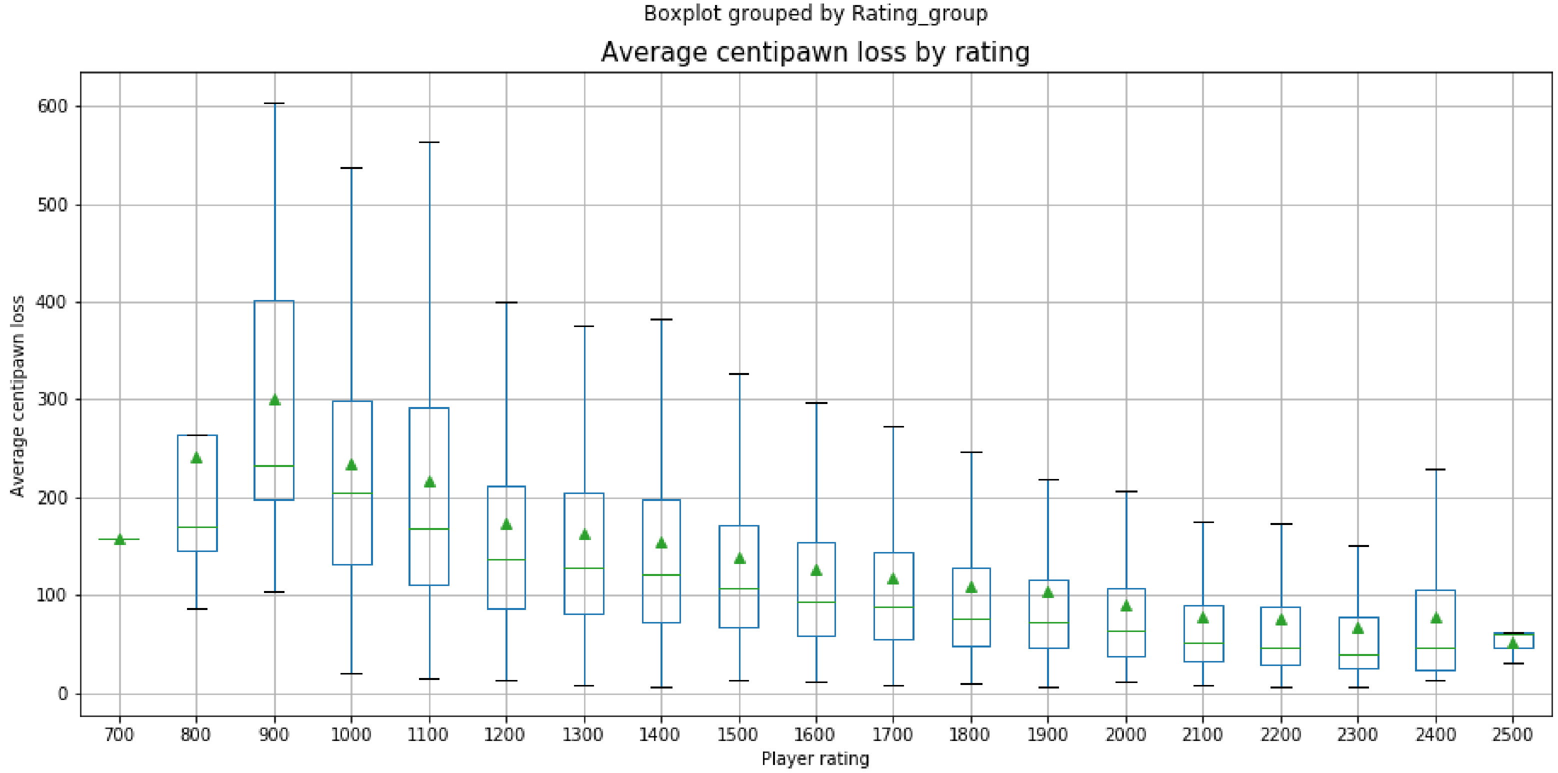
Centipawn loss vs player rating

- There is a clear tendency that CPL decreases with increasing rating.
- The variation is large and increases with lower ratings, indicating less predictable performance.
- The green line is a visual estimate of the main orientation of the plot.



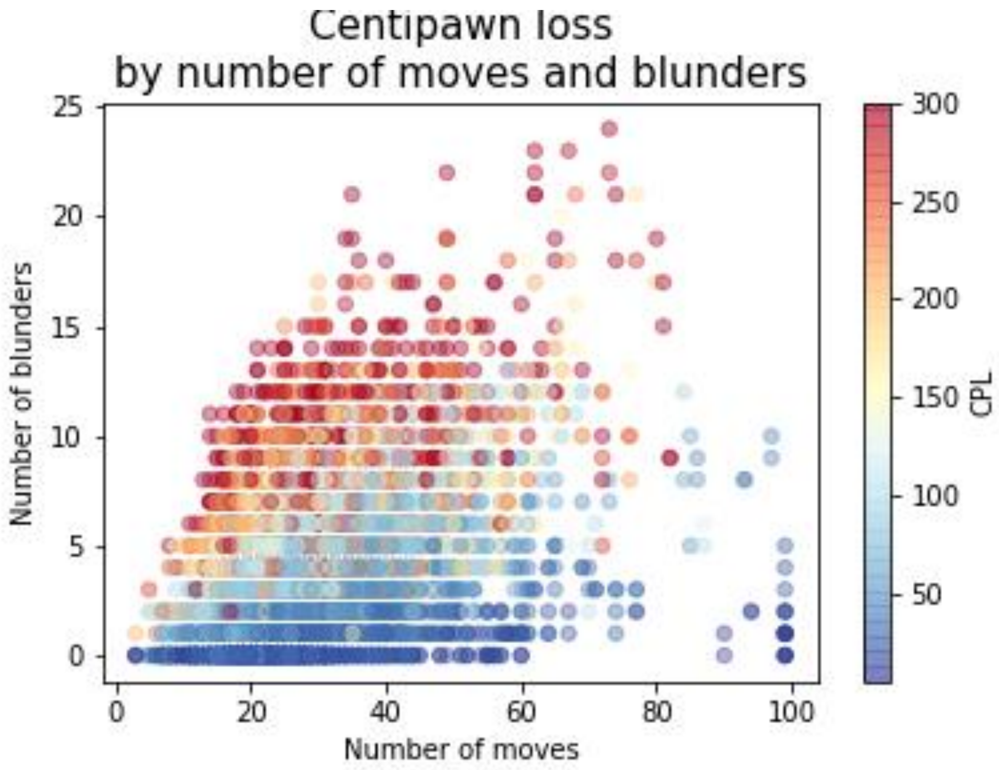
Centipawn loss vs player rating

- The boxplots verify that CPL decreases with player rating
- The variation is quite large
- The overlapping boxes indicate that there is no statistically significant difference between adjacent rating groups.



Centipawn loss and blunders versus number of moves

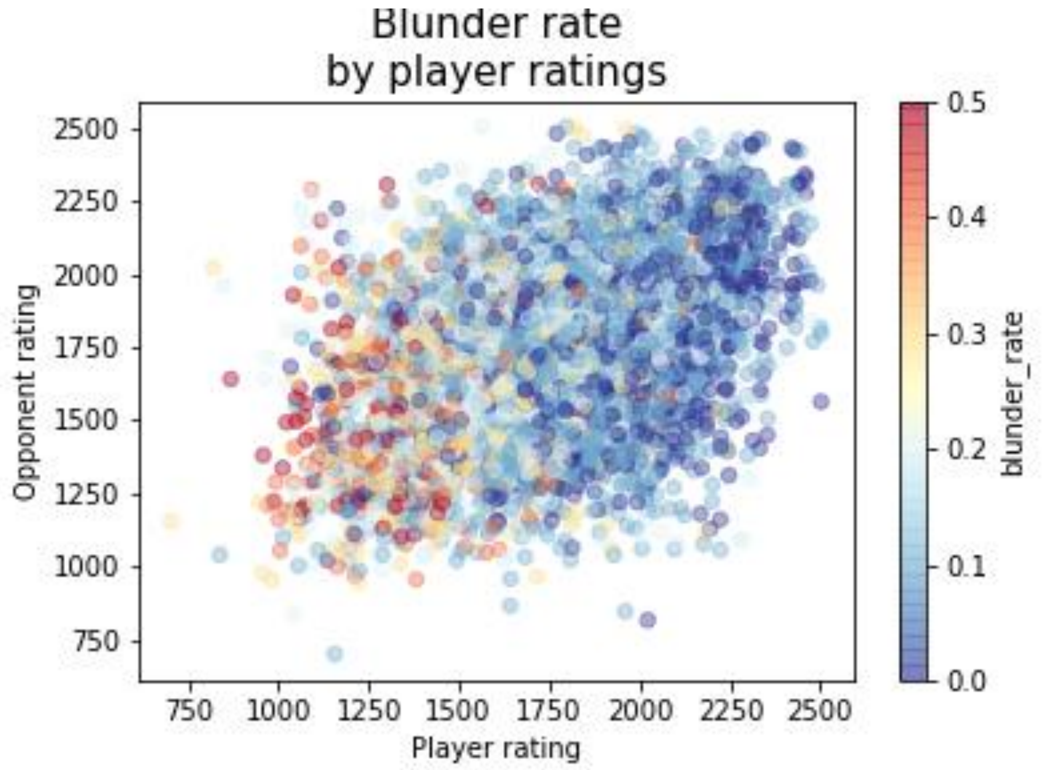
- This plot gives an expected result
- The number of blunders tend to increase with number of moves
- The CPL scores increase with number of blunders



Blunder rate versus rating

This plot gives a partial verification of the apriori assumptions:

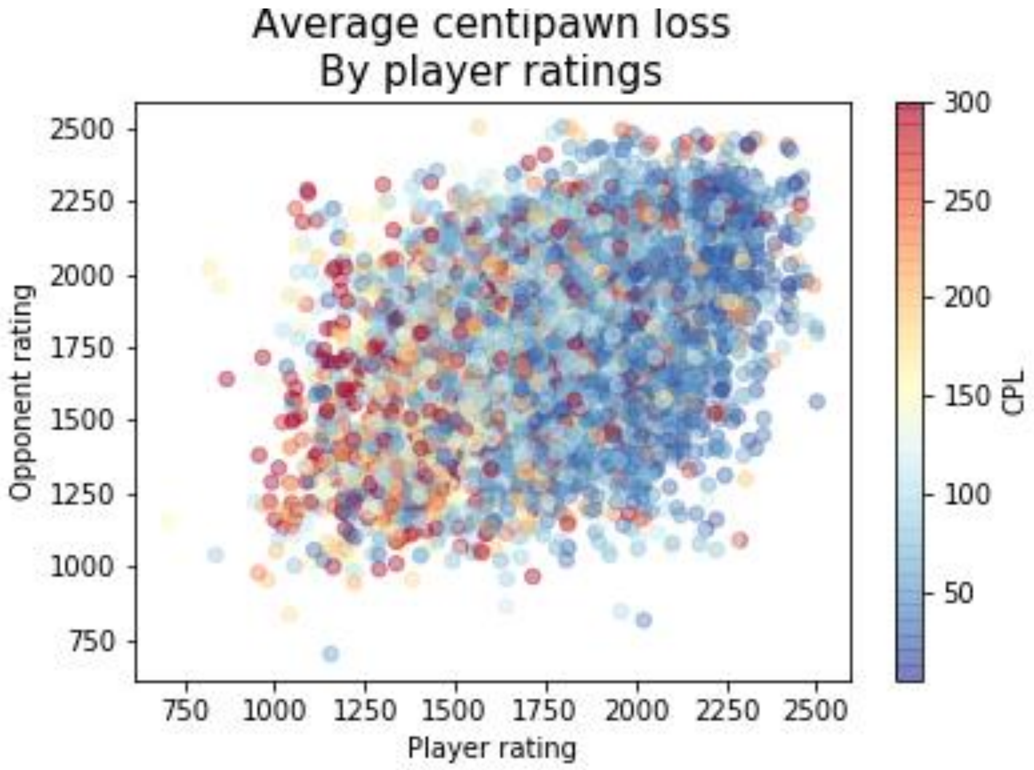
- The blunder rate is higher for lower rated players
- Players in the intermediate range (~1750-2000) seem to make more mistakes against higher rated players



Centipawn loss versus ratings

This plot gives a similar result as the previous one:

- CPL scores are higher in the lower rating ranges
- CPL scores in the intermediate range are lower for stronger opponents
- Occasional low scores are ”scattered” all over the plot, indicating that even strong players make big mistakes and can have very high CPL scores.
- High blunder rates are not ”scattered” in the same way, which indicates that stronger players can make occasional (sometimes huge) mistakes, but the frequency of those mistakes is lower.



Regression model

Based on the analysis above, a regression model was created for the chosen variables. Some variables were exluded due to multicolinearity.

The final regression model is given below. All parameters are significant ($p<0.001$).

$$Rating_{\text{pred}} = 1655 - 0.20 \cdot CPL - 0.45 \cdot RatingDiff + 8.55 \cdot nmoves - 22 \cdot nblunders + \varepsilon$$

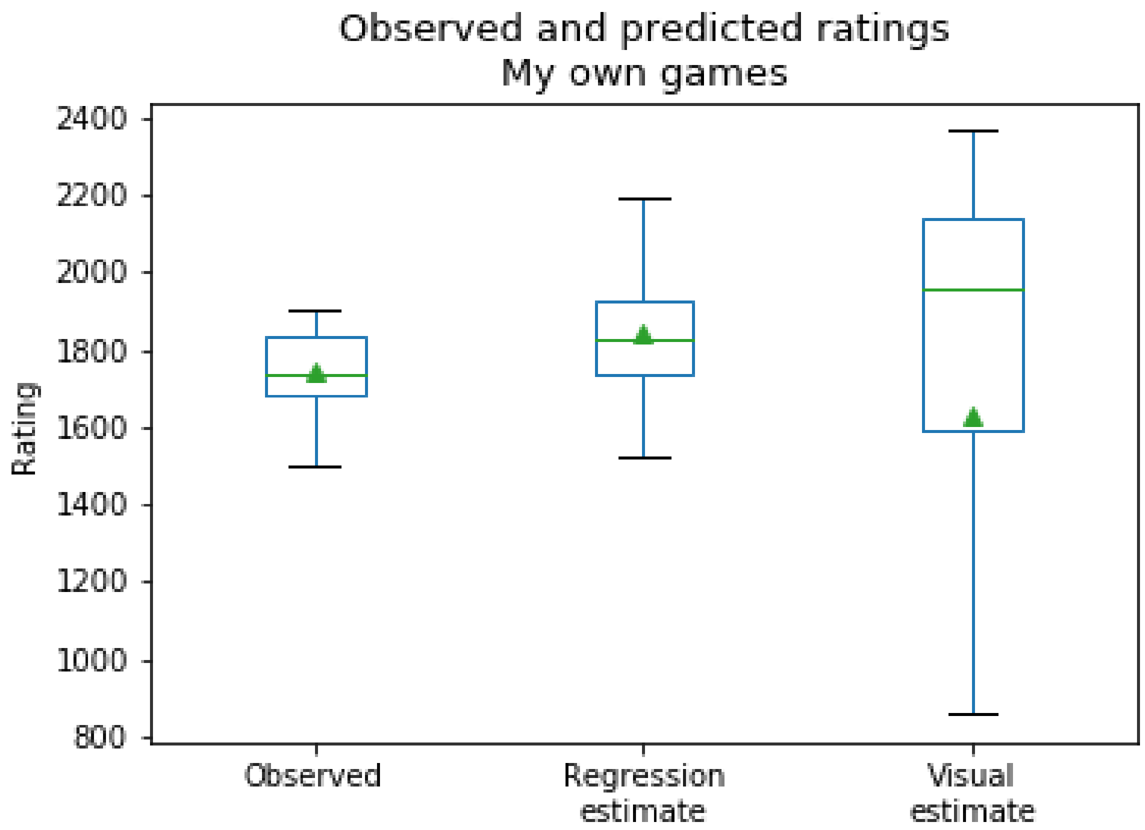
$$\varepsilon \sim N(0, 218)$$

The residuals have a symmetrical distribution that aligns with the assumption of normality. The standard deviation is 218 rating points.

This means that the model can predict playing strength with an accuracy of about ± 400 rating points. It is therefore not very useful for individual games.

Validation of the model

- The regression model was tested for a single player (myself)
- About 30 games were used
- Comparison was made for a prediction based solely on CPL
- The regression model gives a somewhat higher rating estimate compared to observed values
- The predicted values have a standard deviation of 166, indicating that the predicted value has a precision of about 60 rating points (95% CI).



Conclusions

This project has shown that it is possible to estimate playing strength from the information in pg n files. However, estimating playing strength from a single game is not reasonable. Neither is usi ng centipawn loss (CPL) alone as a predictor. This project has resulted in a regression model that seems to be capable of estimating playing strength. However, the model is fairly complex for eve ryday use.