Methods for Data Gathering
- Collect games at every level
- Pull as many and as recent games as possible
  - https://database.lichess.org/
- Looking to pull chess.com games as ratings are more standard
  - Would need to account for rating inflation on lichess

Methods for Data Processing
- Get text list of game moves (one of the columns)
- For every game, calculate:
  - The amount of pieces it was taken for
  - the amount of squares each type of piece moved
  - the amount of checks it made
  - If it was part of the checkmate (boolean)
- Segment across levels, compare differences

Methods for Analytics and Modeling (most important)
- Want to start off with unsupervised learning - see if I can find a way to create an arbitrary value for a piece without using the old ones as an anchor
  - Using logistic regression to create log odds for each piece
  - Inspiration: https://www.r-bloggers.com/2015/06/big-data-and-chess-what-are-the-predictive-point-values-of-chess-pieces/
- Another unsupervised learning method would be looking into association - if groups of pieces or similar patterns occurred with the data known above that makes a piece or group of pieces more valuable at one level than another
- Could then go into some basic supervised learning with regressions
  - Seeing what contributes to winning at different levels

Is there an obvious outcome / classification / model output that you can describe
- The model output is pretty obvious, I will create a table / data frame that estimates the piece value for each general piece for each level of play
  - Pieces: Pawn, knight, bishop, rook, queen, king?
  - Levels: Novice (<800), Beginner (800-1099), Intermediate (1100-1399), Intermediate2 (1400-1699), Advanced (1700-1999), Expert (2000-2299)
- **Example** at Beginner a Bishop is estimated at 2.5 while at Intermediate2 it's 3.7

How will Methods be documented
- Key Steps
  - Write down what the model knows about the games (are arbitrary values put in)
  - Analyze the games via the method
  - Output what the method takes from the game, its assumptions

References to Literature in techniques or methods
- Papers / Articles of Inspiration
  - https://www.r-bloggers.com/2015/06/big-data-and-chess-what-are-the-predictive-point-values-of-chess-pieces/
- Chess Programming Piece Value
  - https://www.chessprogramming.org/Point_Value#Theoretical_Attempt

What is New about your research
- There has been other research related but two key differences
  - Others have been a lot more math focused
  - No segmentation / differentiation by level of play, mostly use grandmaster play
- Different perspectives to evaluation of current stance
  - Where a sacrifice might make sense to an evaluation or the computer's best move, maybe keeping a certain piece might have more value for a certain level of player because they use that piece better than another piece because of their skill level and not necessarily because of best move possible

What is the Contribution - to the field, to policy , to research ,
- Help students at different levels see what pieces they undervalue - how levels above them use them
  - Often many younger / less experienced players just watch grandmasters, where instead they could be watching and understanding the level above them and could be growing gradually

Are there biases or blind spots
- Generalizing ratings as a type of player
  - Aka all 1000 rated players play the same
  - Would want to create an algorithm that just needs ~50 games from a single player to make confident values for that specific player
- Would want chess.com ratings, because lichess ratings are biased