Milestone 4

Levin Neumann

2024-10-31

0. Introduction

I would like to examine what strategies in chess lead to victory. Although technically a solved game, to the average player there are countless unique board states and games to be played, so what can you do to win? Here are the helper functions used.

```r
# Convert square ID to x, y
square_to_pos <- function(square) {
  # Calculate the x and y positions
  x <- (square %% 8) + 1
  y <- (square %/% 8) + 1
  return(c(x, y))
}
# Rotate the board for black
flip_square_pos <- function(square) {
  # Flip
  square <- 63 - square
  # Calculate the x and y positions
  x <- (square %% 8) + 1
  y <- (square %/% 8) + 1
  return(c(x, y))
}

# Convert square name to x, y
move_to_pos <- function(square) {
  # Column index (index at 1)
  x <- utf8ToInt(substr(square, 1, 1)) - utf8ToInt("a") + 1
  # Row number as integer
  y <- as.integer(substr(square, 2, 2))
  return(c(x, y))
}
flip_move_to_pos <- function(square) {
  pos <- move_to_pos(square)
  x <- 9 - pos[1]
  y <- 9 - pos[2]
  return(c(x, y))
}
```

1. White or Black

```
games <- read.csv("games.csv")

# Summarize outcomes to make out favorite graphic!!
winner_summary <- games %>%
  group_by(winner) %>%
  summarise(count = n()) %>%
  mutate(percentage = count / sum(count) * 100)

# Plot
ggplot(winner_summary, aes(x = 2, y = percentage, fill = winner)) +
  geom_bar(stat = "identity", width = 1, color = "black") +
  coord_polar(theta = "y") +
  xlim(1, 2.5) +
  labs(title = "Which Color Should You Play?",
       subtitle = "Assuming you are the average player",
       caption = "Across rated and unrated games.",
       fill = "Outcome") +
  theme_void() +
  scale_fill_manual(values = c("white" = "white",
                               "draw"= "grey",
                               "black" = "black")) +
  theme(legend.position = "right")
```


Which Color Should You Play?

Assuming you are the average player
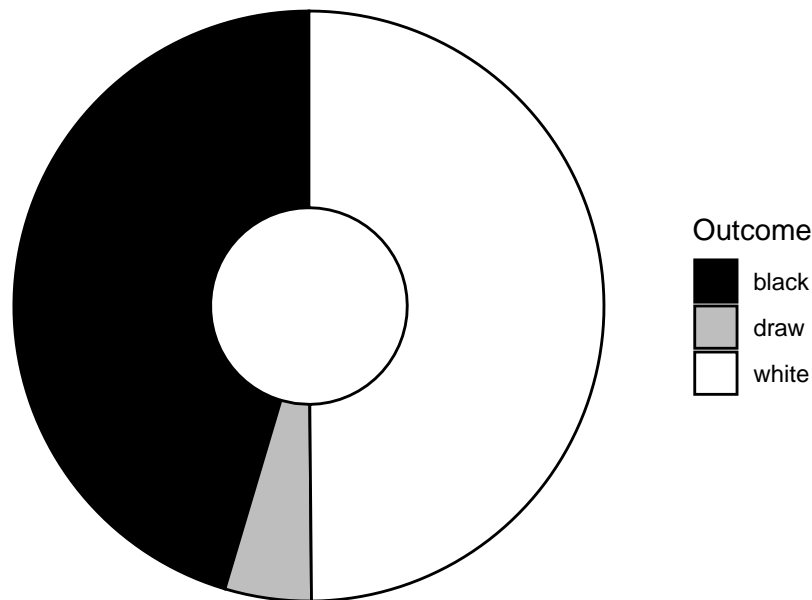
Across rated and unrated games.

This was surprising for me to see, as I thought that it would be an even split. This is suggesting that playing White will actually give you an advantage to win.

2. E4

```r
# Get data
openings <- read.csv("./UseableCSV/openings.csv")

# Change moves into positions
# e4 -> (5, 4)
openings <- openings %>%
  rowwise() %>%
  mutate(pos = list(move_to_pos(first_move))) %>%
  unnest_wider(pos, names_sep = "_") %>%
  rename(x = pos_1, y = pos_2)

# Plot
ggplot(openings, aes(x = x, y = y, fill=count)) +
  geom_tile(width = 0.9, height = 0.9, color = "black") +
  scale_x_continuous(breaks = 1:8, labels = c("a", "b", "c", "d", "e", "f", "g", "h"), limits = c(1, 8))
  scale_y_continuous(breaks = 1:8, limits = c(1, 8)) +
  scale_fill_viridis_c(trans = "log") +
  coord_fixed() +
  labs(title = paste("Who Wins with Opening Moves"),
       subtitle = "Split on who won the game",
       caption = "The count is using a log scale to be able to see lower values.",
       fill = "Count",
       x =  "Column", y = "Row") +
  theme_bw() +
  facet_wrap(~ outcome)
```
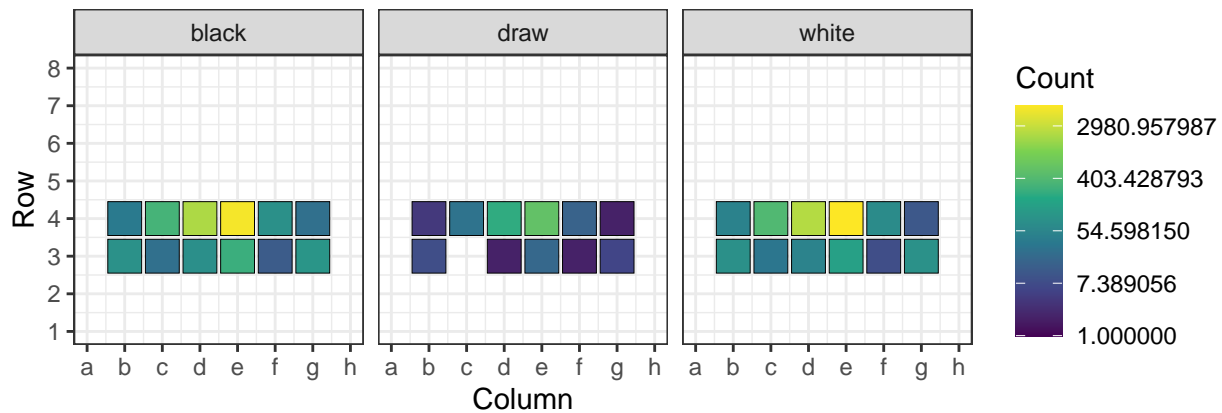
Who Wins with Opening Moves
### Split on who won the game

The count is using a log scale to be able to see lower values.

Rather disappointingly there is not a clear way too see if you will win from the first move. This is however not the most surprising given that if there was it would have likely been found out before. However that being said, as White if you play with a good pawn structure or go for the center squares it seems you will be more likely to win.

3. Castling

```
castle_w <- read.csv("./UseableCSV/castle_winners.csv")


ggplot(castle_w, aes(x = outcome, y = count, fill = castling_player)) +
    geom_bar(stat = "identity", color="black", position = position_dodge(width = 0.7), width = 0.6) +
    coord_flip() +
    labs(
        title = "Castling to Victory",
        subtitle = "Who wins from castling",
        caption = "Outcome of games when a player castled",
        x = "Game Outcome",
        y = "Count of Castling",
        fill = "Player"
    ) +
    theme_bw() +
    scale_fill_manual(values = c("white" = "white", "black" = "black"))
```

Castling to Victory
Who wins from castling

Count of Castling

Outcome of games when a player castled
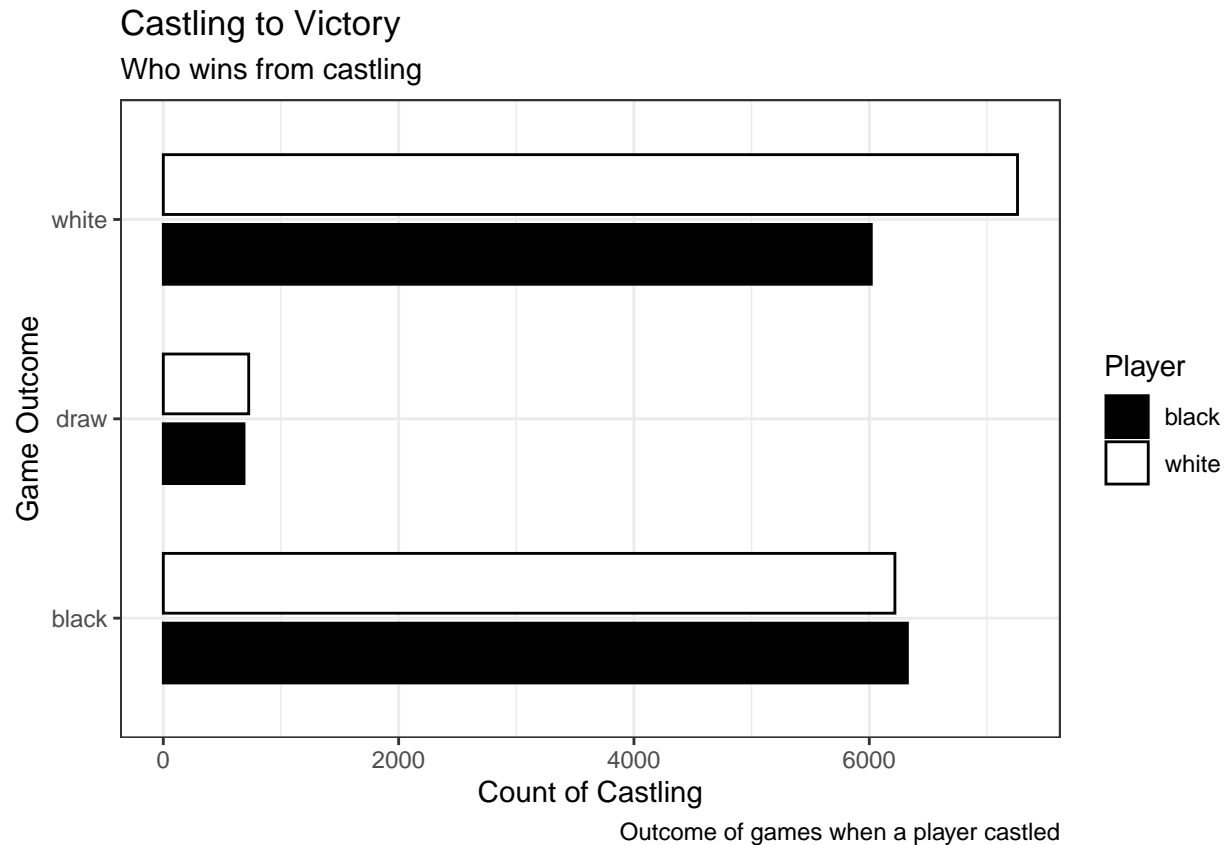
Not only does White castle more, but it seems that White castling will usually either be beneficial or near negligible when comparing how often players castled when Black won.
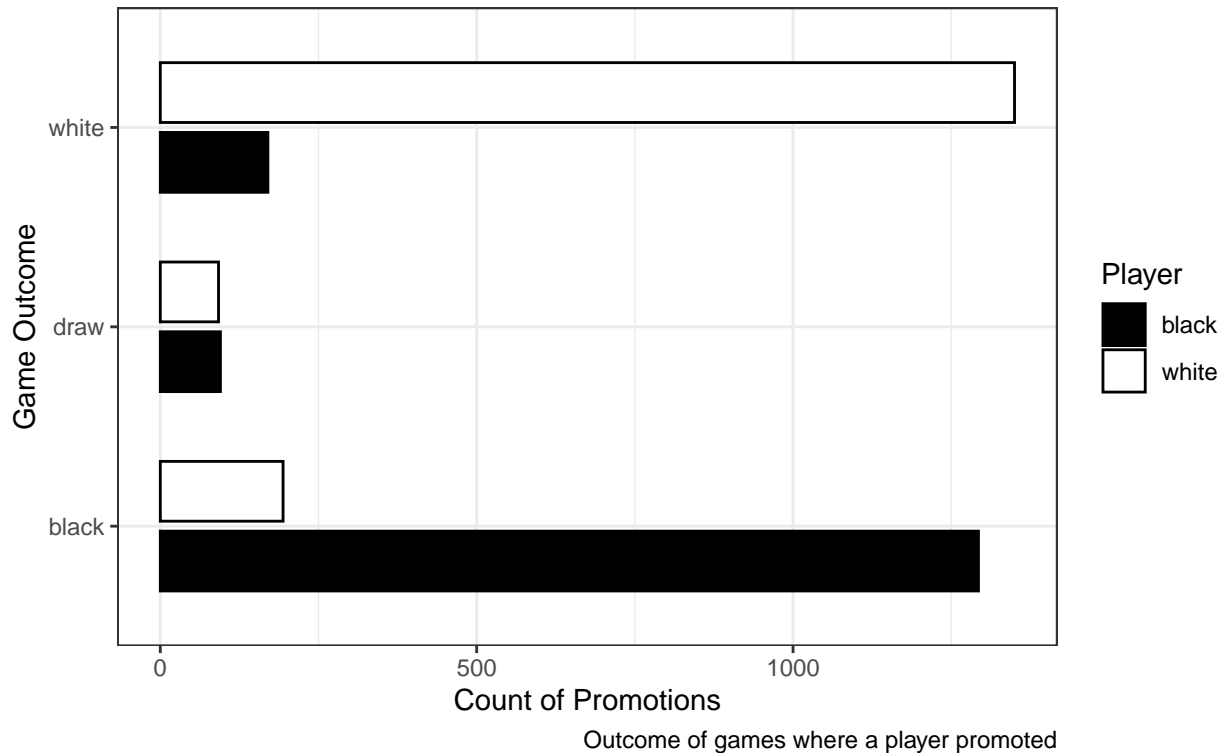
4. Promotions

```
promotion_w <- read.csv("./UseableCSV/promotion_wins.csv")


ggplot(promotion_w, aes(x = outcome, y = count, fill = promotes)) +
    geom_bar(stat = "identity", color="black", position = position_dodge(width = 0.7), width = 0.6) +
    coord_flip() +
    labs(
        title = "Promote Your Winrate?",
        subtitle = "Do promotions actually win games? (Yes, yes they do)",
        caption = "Outcome of games where a player promoted",
        x = "Game Outcome",
        y = "Count of Promotions",
        fill = "Player"
    ) +
    theme_bw() +
    scale_fill_manual(values = c("white" = "white", "black" = "black"))
```

Promote Your Winrate?

Do promotions actually win games? (Yes, yes they do)

Outcome of games where a player promoted

Promotions are generally considered one of the most powerful actions you can take. This is clearly seen here. When players promote it makes it significantly more likely to win.
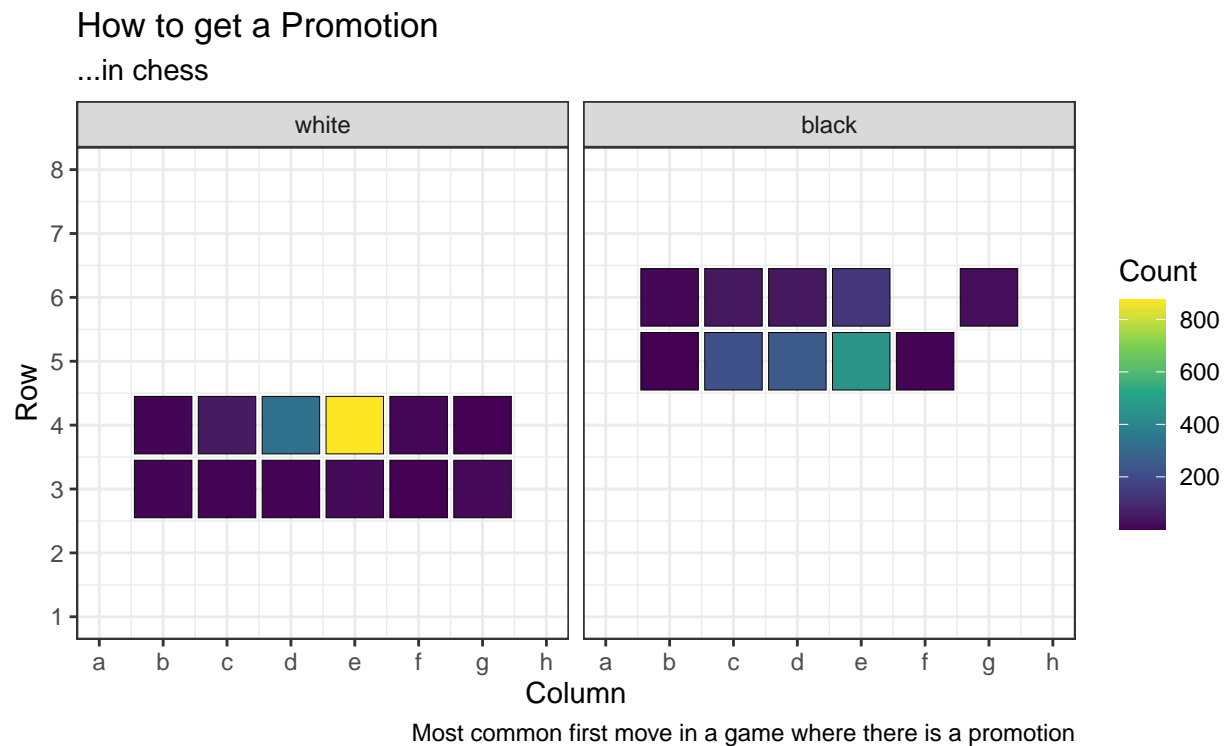
5. How to promote

```r
# Get data
promote_o <- read.csv("./UseableCSV/promote_opening.csv")

# Change moves into positions, and flip board
# e4 -> (5, 4)
promote_o <- promote_o %>%
  rowwise() %>%
  mutate(pos = if_else(player == "black", list(move_to_pos(first_move)), list(move_to_pos(first_move)))
  unnest_wider(pos, names_sep = "_") %>%
  rename(x = pos_1, y = pos_2)
promote_o$player <- factor(promote_o$player, levels = c("white", "black"))

# Create a custom y-axis label function
y_labels <- function(player) {
  if (player == "black") {
    return(rev(1:8))
  } else {
    return(1:8)
  }
}
```

```
# Plot
ggplot(promote_o, aes(x = x, y = y, fill=count)) +
  geom_tile(width = 0.9, height = 0.9, color = "black") +
  scale_x_continuous(breaks = 1:8, labels = c("a", "b", "c", "d", "e", "f", "g", "h"), limits = c(1, 8))
  scale_y_continuous(breaks = 1:8, limits = c(1, 8)) +
  scale_fill_viridis_c() +
  coord_fixed() +
  labs(title = paste("How to get a Promotion"),
       subtitle = "...in chess",
       caption = "Most common first move in a game where there is a promotion",
       fill = "Count",
       x =  "Column", y = "Row") +
  theme_bw() +
  facet_wrap(~ player)
```


How to get a Promotion
...in chess

Most common first move in a game where there is a promotion

Since we established that promotions are very good, we want to know how to promote. Notable, as E4 is the most common move in general, this leads to E4 being very bright on this graphic. That being said it seems like the E file does lead to more promotions.