

Model Proposal

Ben Moolman, Craig Orman, Ethan Pross

What I want is a scatter plot. Y is FGA, X is opposing team name, and each dot represents one of the games played against them.

```
df = distinct(starting_dat[,c(2,3, 4)])

# game_registry = df %>%
#   group_by(GAME_ID) %>%
#   arrange(TEAM_ID, .by_group = TRUE) %>% # Optional: keeps order consistent
#   mutate(team_slot = paste0("TEAM", row_number(), "_ID")) %>%
#   pivot_wider(
#     names_from = team_slot,
#     values_from = TEAM_ID
#   )

game_registry <- df %>%
  group_by(GAME_ID) %>%
  arrange(TEAM_ID, .by_group = TRUE) %>%
  mutate(slot = paste0("TEAM", row_number())) %>%
  pivot_wider(
    names_from = slot,
    values_from = c(TEAM_ID, TEAM_ABBREVIATION),
    names_sep = "_"
  )

colnames(game_registry) <- c("Game_ID", "Team_1_ID", "Team_2_ID", "Team_1_Abbrev", "Team_2_Abbrev")

starting_dat$Opposing_Team_Name_ID = NA
starting_dat$Opposing_Team_Name = NA

for (i in 1:nrow(starting_dat)) {
  for (j in 1:nrow(game_registry)) {

    if (starting_dat$GAME_ID[i] == game_registry$Game_ID[j]){

      if (starting_dat$TEAM_ID[i] == game_registry$Team_1_ID[j]) {
        starting_dat$Opposing_Team_Name_ID[i] = game_registry$Team_2_ID[j]
        starting_dat$Opposing_Team_Name[i] = game_registry$Team_2_Abbrev[j]
      }
      if (starting_dat$TEAM_ID[i] == game_registry$Team_2_ID[j]) {
        starting_dat$Opposing_Team_Name_ID[i] = game_registry$Team_1_ID[j]
        starting_dat$Opposing_Team_Name[i] = game_registry$Team_1_Abbrev[j]
      }
    }
  }
}
```

```

    }
  }
}

lebron_dat = starting_dat[starting_dat$PLAYER_ID %in% 2544, ] # lebron 2544 and steph curry 201939 anth
anthony_dat = starting_dat[starting_dat$PLAYER_ID %in% 203076, ]

```

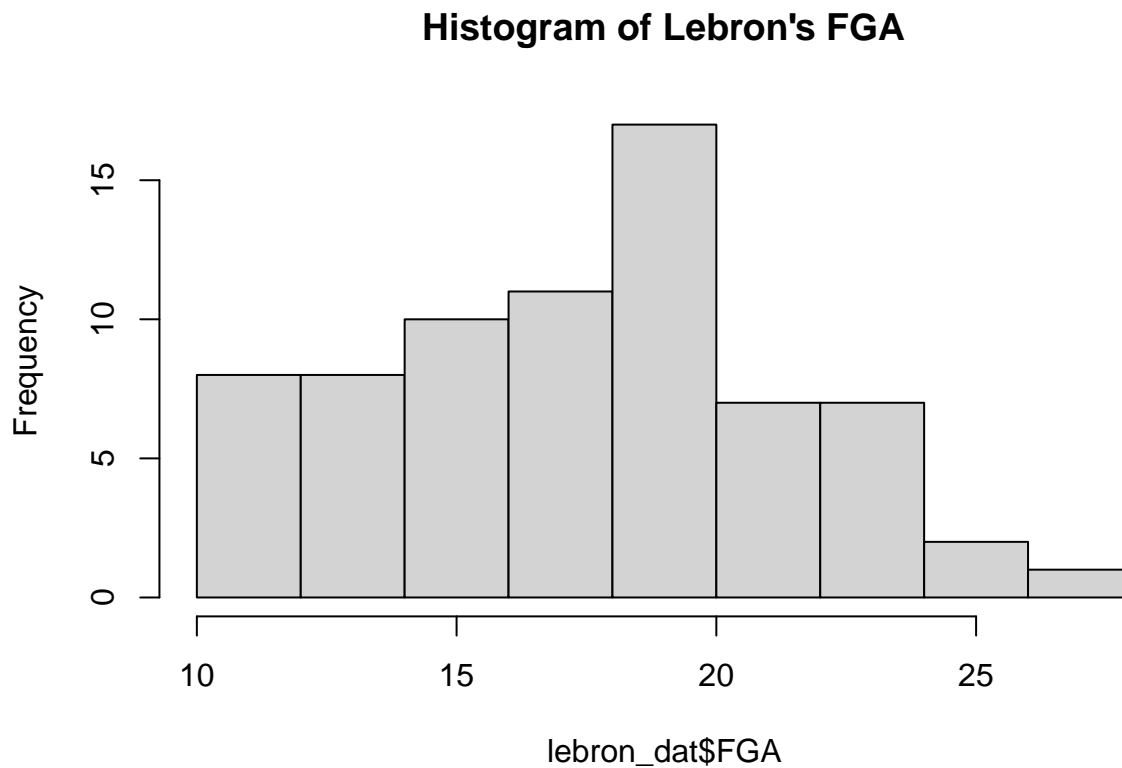
Examining FGA and FGM

```

team_ids = lebron_dat[lebron_dat$PLAYER_ID == 2544,c(2,3)]

hist(lebron_dat$FGA, main="Histogram of Lebron's FGA")

```



```

# hist(rpois(1000, 10))

```

```

library(ggplot2)
library(forcats)
library(forcats)

lebron_dat = starting_dat[starting_dat$PLAYER_ID %in% 2544, ] # lebron 2544 and steph curry 201939
lebron_dat <- lebron_dat %>%
  mutate(Opposing_Team_Name = fct_reorder(as.factor(Opposing_Team_Name), FGA, .fun = mean))
mean_fga = mean(lebron_dat$FGA)

```

```
sd_fga = sd(lebron_dat$FGA)

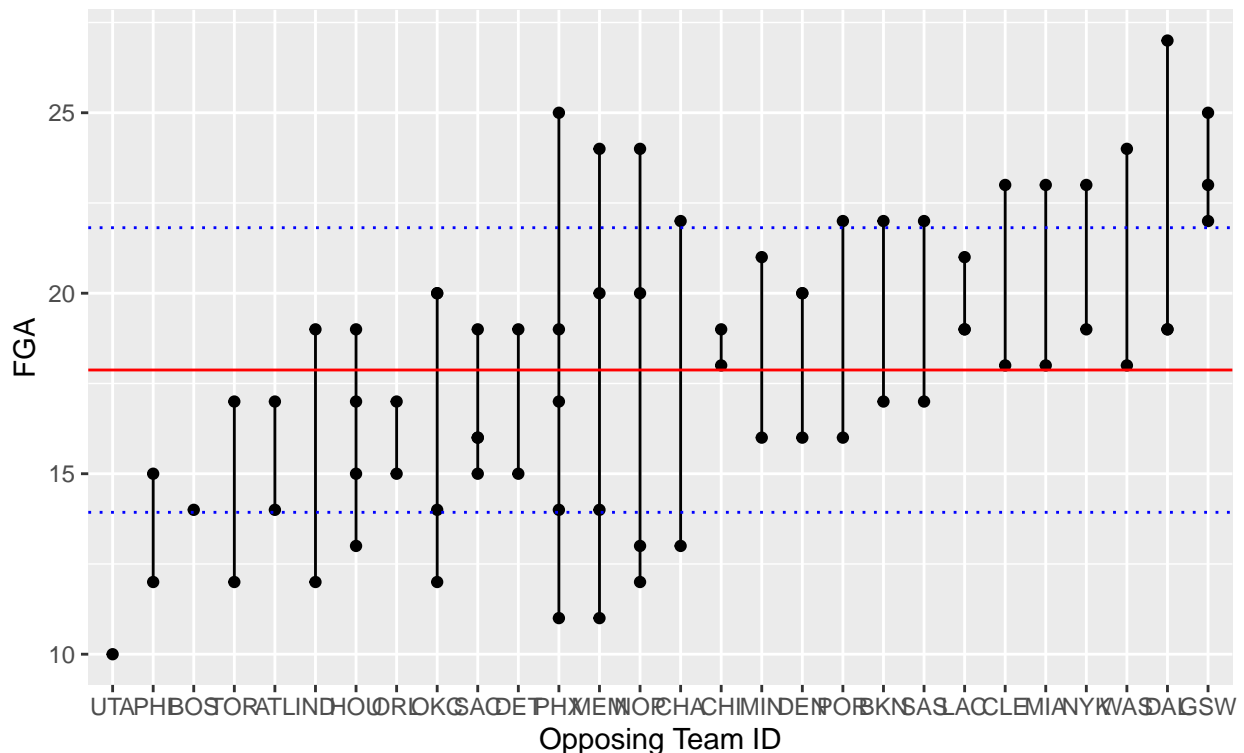
df_lines <- lebron_dat %>%
  group_by(Opposing_Team_Name) %>%
  arrange(FGA) %>%
  mutate(index = row_number()) # create a pseudo-x to keep line ordering

ggplot(df_lines, aes(x = FGA, y = Opposing_Team_Name, group = Opposing_Team_Name)) +
  geom_point(aes(x = Opposing_Team_Name, y = FGA)) +
  geom_line(aes(x = Opposing_Team_Name, y = FGA, group = Opposing_Team_Name), orientation = "FGA") +
  labs(
    title = "Lebron FGA group variation plot",
    subtitle = "Data is only for Lebron, includes mean in red and 1 sd in blue",
    x = "Opposing Team ID",
    y = "FGA"
  ) +
  geom_hline(yintercept = mean_fga, color = "red", linetype = "solid", size = 0.5) +
  geom_hline(yintercept = mean_fga - sd_fga, color = "blue", linetype = "dotted", size = 0.5) +
  geom_hline(yintercept = mean_fga + sd_fga, color = "blue", linetype = "dotted", size = 0.5)

## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Lebron FGA group variation plot

Data is only for Lebron, includes mean in red and 1 sd in blue



```

# And now for Steph
steph_dat = starting_dat[starting_dat$PLAYER_ID == 201939, ] # lebron 2544 and steph curry 201939

steph_dat$Opposing_Team_Name <- factor(steph_dat$Opposing_Team_Name, levels = levels(lebron_dat$Opposing_Team_Name))
mean_fga = mean(steph_dat$FGA)
sd_fga = sd(steph_dat$FGA)

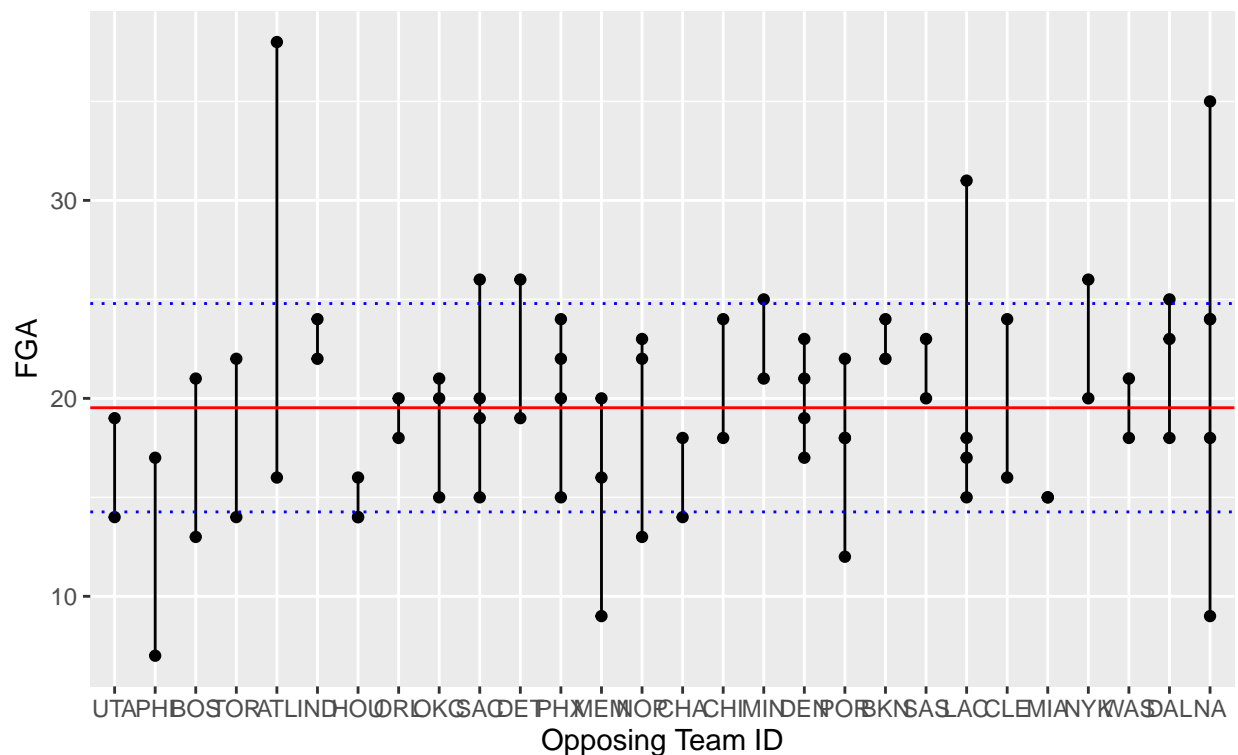
df_lines <- stp_dat %>%
  group_by(Opposing_Team_Name) %>%
  arrange(FGA) %>%
  mutate(index = row_number()) # create a pseudo-x to keep line ordering

ggplot(df_lines, aes(x = FGA, y = Opposing_Team_Name, group = Opposing_Team_Name)) +
  geom_point(aes(x = Opposing_Team_Name, y = FGA)) +
  geom_line(aes(x = Opposing_Team_Name, y = FGA, group = Opposing_Team_Name), orientation = "FGA") +
  labs(
    title = "Steph Curry FGA group variation plot",
    subtitle = "Data is only for Steph Curry, includes mean in red and 1 sd in blue",
    x = "Opposing Team ID",
    y = "FGA"
  ) +
  geom_hline(yintercept = mean_fga, color = "red", linetype = "solid", size = 0.5) +
  geom_hline(yintercept = mean_fga - sd_fga, color = "blue", linetype = "dotted", size = 0.5) +
  geom_hline(yintercept = mean_fga + sd_fga, color = "blue", linetype = "dotted", size = 0.5)

```

Steph Curry FGA group variation plot

Data is only for Steph Curry, includes mean in red and 1 sd in blue



```

library(ggplot2)
library(forcats)
library(forcats)

lebron_dat <- lebron_dat %>%
  mutate(Opposing_Team_Name = fct_reorder(as.factor(Opposing_Team_Name), FG_PCT, .fun = mean))
mean_fga = mean(lebron_dat$FG_PCT)
sd_fga = sd(lebron_dat$FG_PCT)

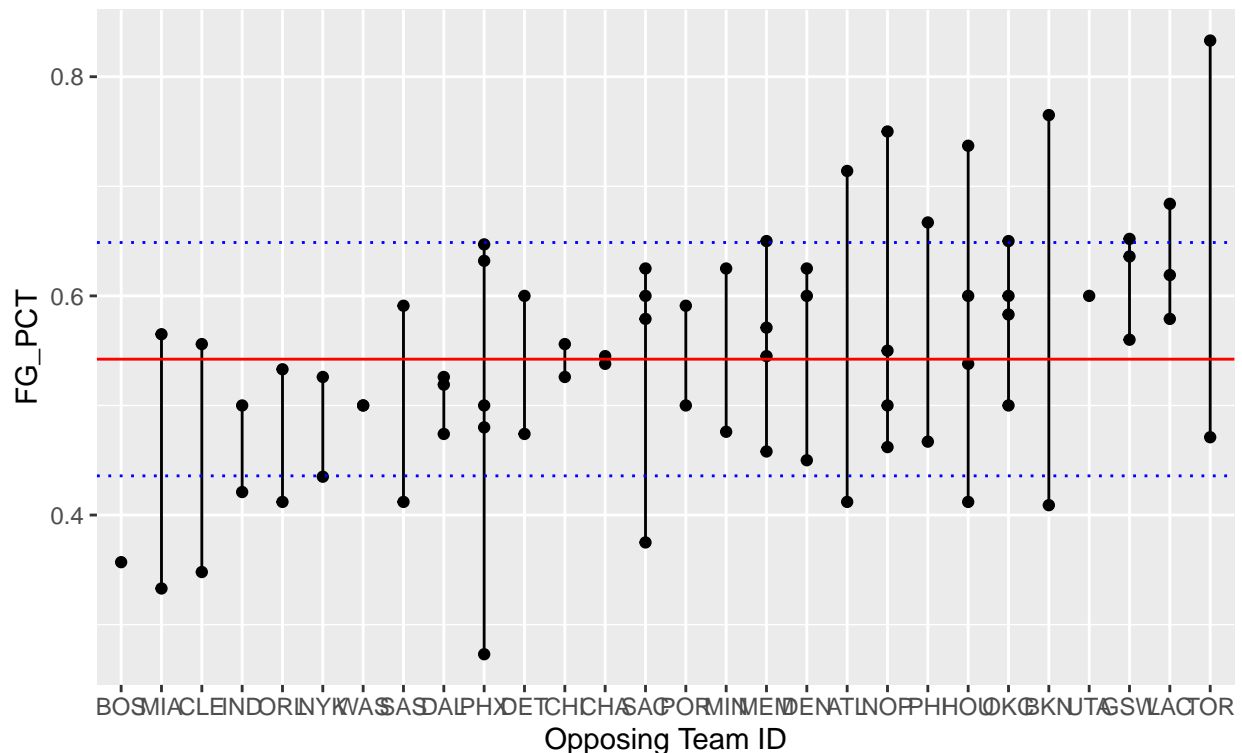
df_lines <- lebron_dat %>%
  group_by(Opposing_Team_Name) %>%
  arrange(FG_PCT) %>%
  mutate(index = row_number()) # create a pseudo-x to keep line ordering

ggplot(df_lines, aes(x = FG_PCT, y = Opposing_Team_Name, group = Opposing_Team_Name)) +
  geom_point(aes(x = Opposing_Team_Name, y = FG_PCT)) +
  geom_line(aes(x = Opposing_Team_Name, y = FG_PCT, group = Opposing_Team_Name), orientation = "FG_PCT",
  labs(
    title = "Lebron FG_PCT group variation plot",
    subtitle = "Data is only for Lebron, includes mean in red and 1 sd in blue",
    x = "Opposing Team ID",
    y = "FG_PCT"
  ) +
  geom_hline(yintercept = mean_fga, color = "red", linetype = "solid", size = 0.5) +
  geom_hline(yintercept = mean_fga - sd_fga, color = "blue", linetype = "dotted", size = 0.5) +
  geom_hline(yintercept = mean_fga + sd_fga, color = "blue", linetype = "dotted", size = 0.5)

```

Lebron FG_PCT group variation plot

Data is only for Lebron, includes mean in red and 1 sd in blue



```
library(ggplot2)
library(forcats)
library(forcats)

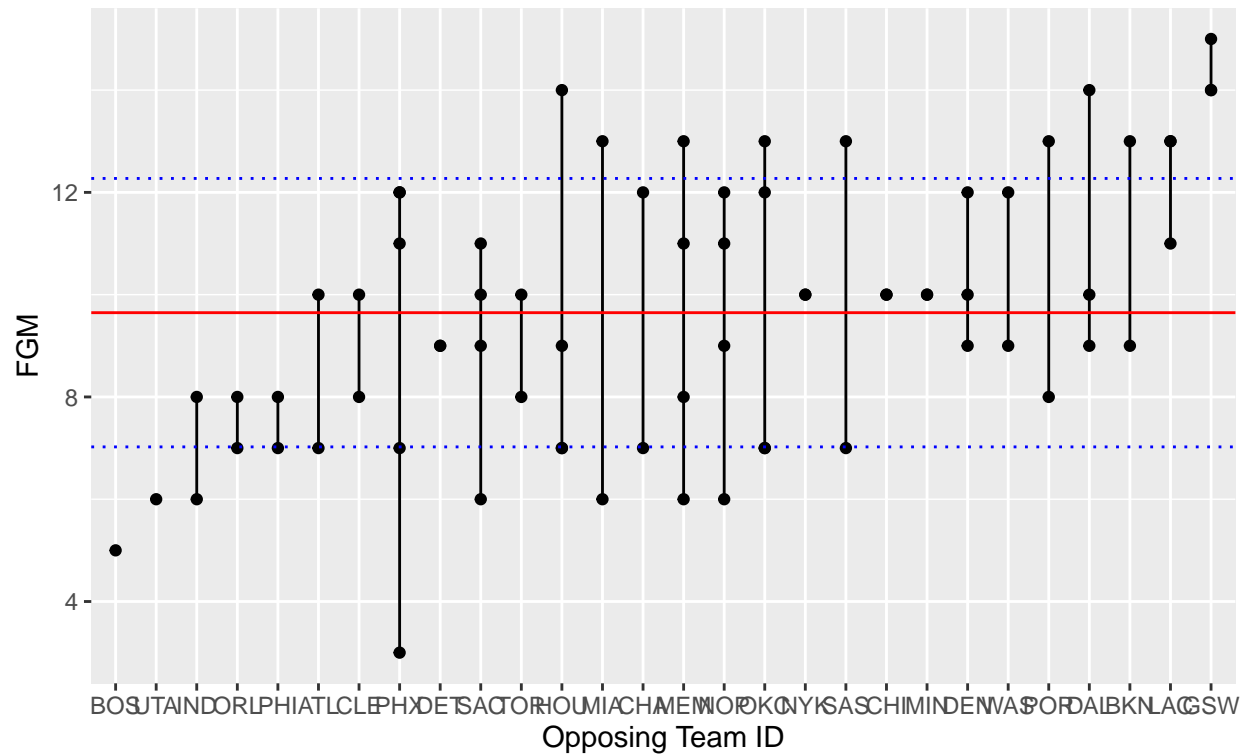
lebron_dat <- lebron_dat %>%
  mutate(Opposing_Team_Name = fct_reorder(as.factor(Opposing_Team_Name), FGM, .fun = mean))
mean_fga = mean(lebron_dat$FGM)
sd_fga = sd(lebron_dat$FGM)

df_lines <- lebron_dat %>%
  group_by(Opposing_Team_Name) %>%
  arrange(FGM) %>%
  mutate(index = row_number()) # create a pseudo-x to keep line ordering

ggplot(df_lines, aes(x = FGM, y = Opposing_Team_Name, group = Opposing_Team_Name)) +
  geom_point(aes(x = Opposing_Team_Name, y = FGM)) +
  geom_line(aes(x = Opposing_Team_Name, y = FGM, group = Opposing_Team_Name), orientation = "FGM") +
  labs(
    title = "Lebron FGM group variation plot",
    subtitle = "Data is only for Lebron, includes mean in red and 1 sd in blue",
    x = "Opposing Team ID",
    y = "FGM"
  ) +
  geom_hline(yintercept = mean_fga, color = "red", linetype = "solid", size = 0.5) +
  geom_hline(yintercept = mean_fga - sd_fga, color = "blue", linetype = "dotted", size = 0.5) +
  geom_hline(yintercept = mean_fga + sd_fga, color = "blue", linetype = "dotted", size = 0.5)
```

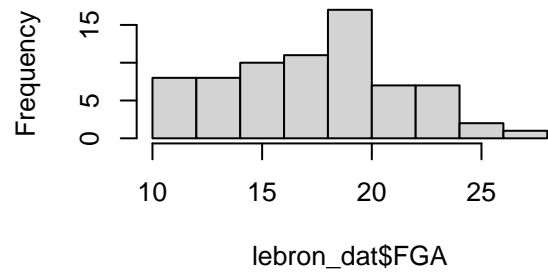
Lebron FGM group variation plot

Data is only for Lebron, includes mean in red and 1 sd in blue



```
par(mfrow=c(2,2))
hist(lebron_dat$FGA)
steph_dat = starting_dat[starting_dat$PLAYER_ID == 201939, ] # lebron 2544 and steph curry 201939
hist(steph_dat$FGA)
```

Histogram of lebron_dat\$FGA



Histogram of steph_dat\$FGA

