

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)])

game_reigstry = 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
  )

starting_dat$Opposing_Team = NA

for (i in 1:nrow(starting_dat)) {
  for (j in 1:nrow(game_reigstry)) {
    if (starting_dat$GAME_ID[i] == game_reigstry$GAME_ID[j]){
      if (starting_dat$TEAM_ID[i] == game_reigstry$TEAM1_ID[j]) {
        starting_dat$Opposing_Team[i] = game_reigstry$TEAM2_ID[j]
      }
      if (starting_dat$TEAM_ID[i] == game_reigstry$TEAM2_ID[j]) {
        starting_dat$Opposing_Team[i] = game_reigstry$TEAM1_ID[j]
      }
    }
  }
}

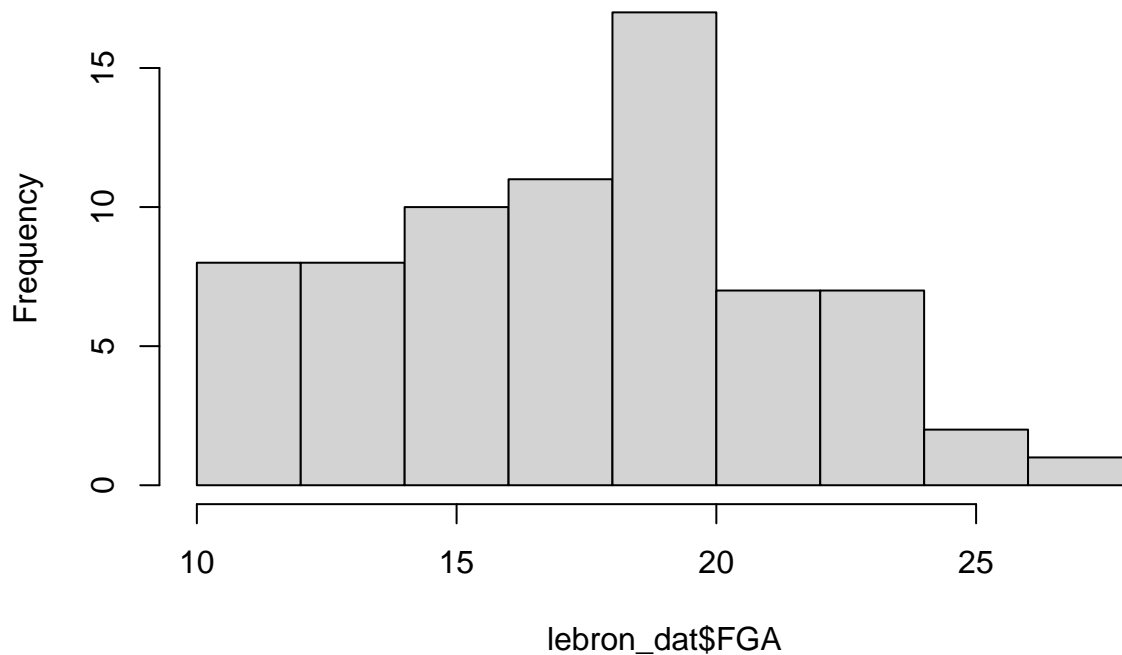
lebron_dat = starting_dat[starting_dat$PLAYER_ID %in% 2544, ] # lebron 2544 and steph curry 201939
```

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")
```

Histogram of Lebron's FGA



```
# hist(rpois(1000, 10))
```

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

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

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

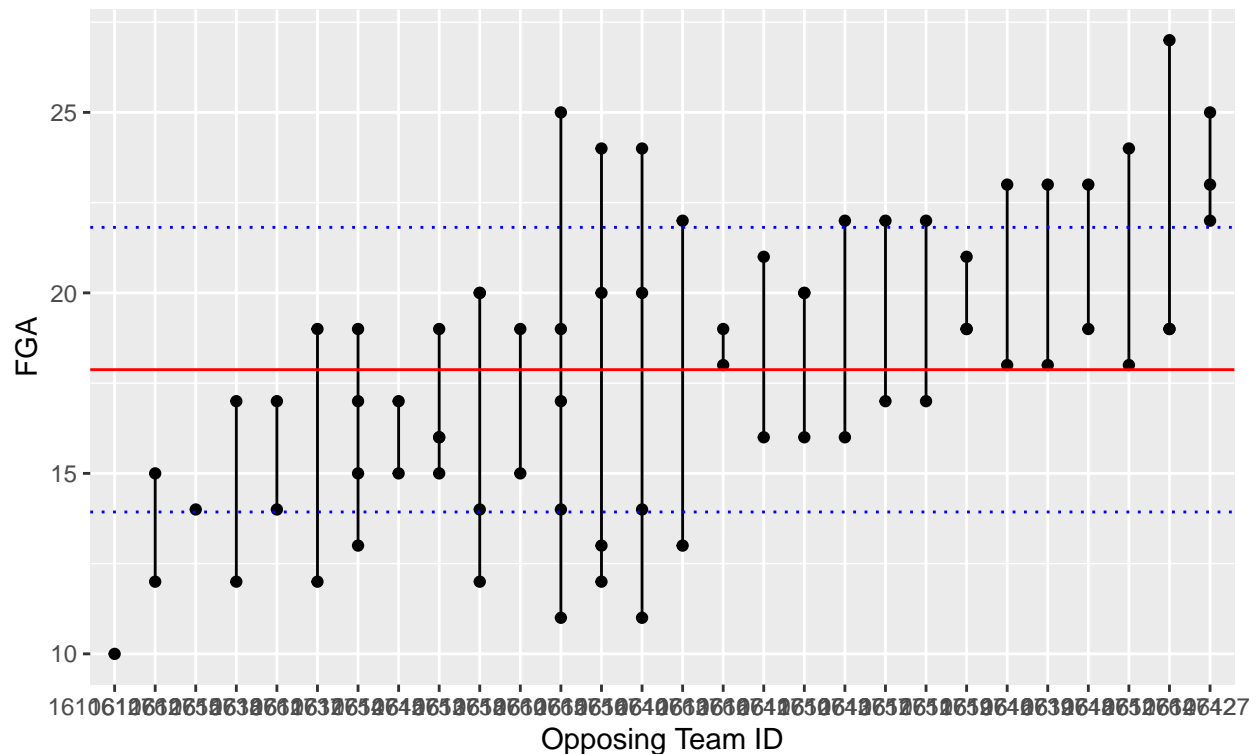
ggplot(df_lines, aes(x = FGA, y = Opposing_Team, group = Opposing_Team)) +
  geom_point(aes(x = Opposing_Team, y = FGA)) +
  geom_line(aes(x = Opposing_Team, y = FGA, group = Opposing_Team), orientation = "FGA") +
  labs(
    title = "Lebron FGA group variation plot",
    subtitle = "Data is only for 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



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

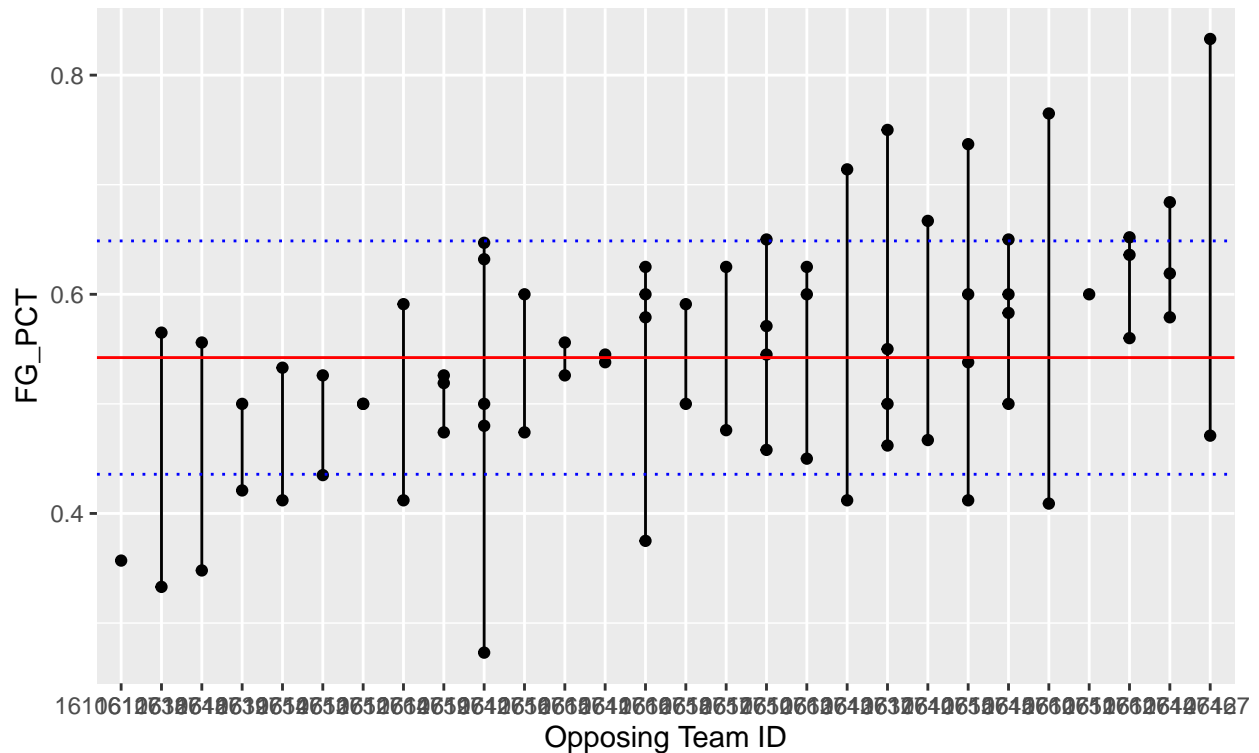
lebron_dat <- lebron_dat %>%
  mutate(Opposing_Team = fct_reorder(as.factor(Opposing_Team), FG_PCT, .fun = mean))
mean_fga = mean(lebron_dat$FG_PCT)
sd_fga = sd(lebron_dat$FG_PCT)

df_lines <- lebron_dat %>%
  group_by(Opposing_Team) %>%
  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, group = Opposing_Team)) +
  geom_point(aes(x = Opposing_Team, y = FG_PCT)) +
  geom_line(aes(x = Opposing_Team, y = FG_PCT, group = Opposing_Team), 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 = fct_reorder(as.factor(Opposing_Team), FGM, .fun = mean))
mean_fga = mean(lebron_dat$FGM)
sd_fga = sd(lebron_dat$FGM)

df_lines <- lebron_dat %>%
  group_by(Opposing_Team) %>%
  arrange(FGM) %>%
```

```
mutate(index = row_number()) # create a pseudo-x to keep line ordering

ggplot(df_lines, aes(x = FGM, y = Opposing_Team, group = Opposing_Team)) +
  geom_point(aes(x = Opposing_Team, y = FGM)) +
  geom_line(aes(x = Opposing_Team, y = FGM, group = Opposing_Team), 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

