

# Assessing Referee Impartiality: The Likelihood of Red Cards for Dark-Skin-Toned Players in Soccer

Ibrahim Uruc Tarim

## Dataset

- playerShort: The short name of the player, useful for identification.
- club: The club for which the player plays, indicating the team context of the player's performance.
- leagueCountry: The country of the league in which the player competes, which could relate to different refereeing styles and biases.
- position: The position the player plays, relevant since certain positions might receive more red cards.
- games: The number of games played, essential for calculating rates like red cards per game.
- yellowCards: The number of yellow cards received, which could indicate a player's disciplinary record.
- redCards: The number of red cards received, a key variable for your analysis.
- rater1 and rater2: Skin tone rating by two different raters, which is central to analyzing potential bias.
- meanIAT: The mean Implicit Association Test score for the referee country, measuring implicit bias levels.

## First approach

```
data <- read.csv("CrowdstormingDataJuly1st.csv")
data$mean_rating <- rowMeans(data[,c("rater1", "rater2")], na.rm = TRUE)
clean_data <- na.omit(data)

avg_red_cards <- clean_data %>%
  group_by(mean_rating) %>%
```

```

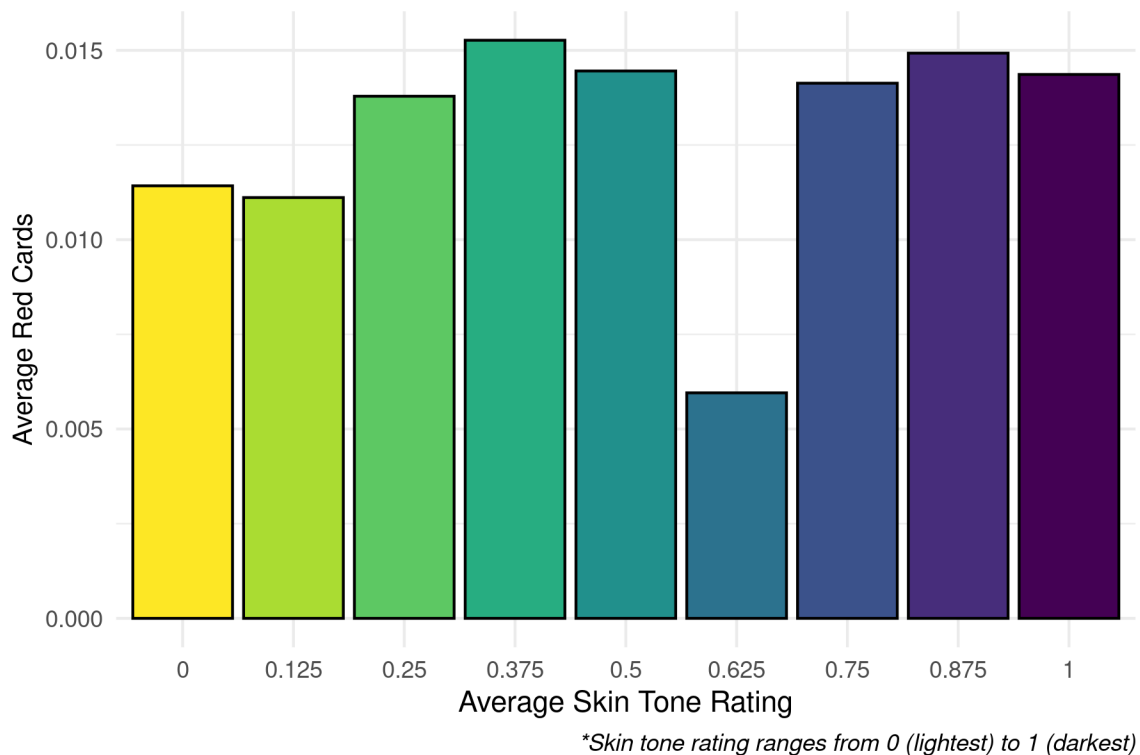
summarise(average_reds = mean(redCards))

ggplot(avg_red_cards, aes(x = as.factor(mean_rating), y = average_reds, fill = as.factor(m
geom_bar(stat = "identity", color = "black") +
scale_fill_viridis(discrete = TRUE, direction = -1) +
theme_minimal() +
theme(legend.position = "none",
      axis.title.y = element_text(size = 10),
      plot.caption = element_text(face = "italic"),
      plot.title = element_text(hjust = 0, size = 14),
      plot.subtitle = element_text(hjust = 0, size = 10),
      plot.title.position = "plot") +
labs(title = "Distribution of Red Cards Across Skin Tone Ratings",
      subtitle = "The trend of red cards does not show a strong linear increase with skin
      caption = "*Skin tone rating ranges from 0 (lightest) to 1 (darkest)",
      x = "Average Skin Tone Rating", y = "Average Red Cards")

```

## Distribution of Red Cards Across Skin Tone Ratings

The trend of red cards does not show a strong linear increase with skin tone rating



```

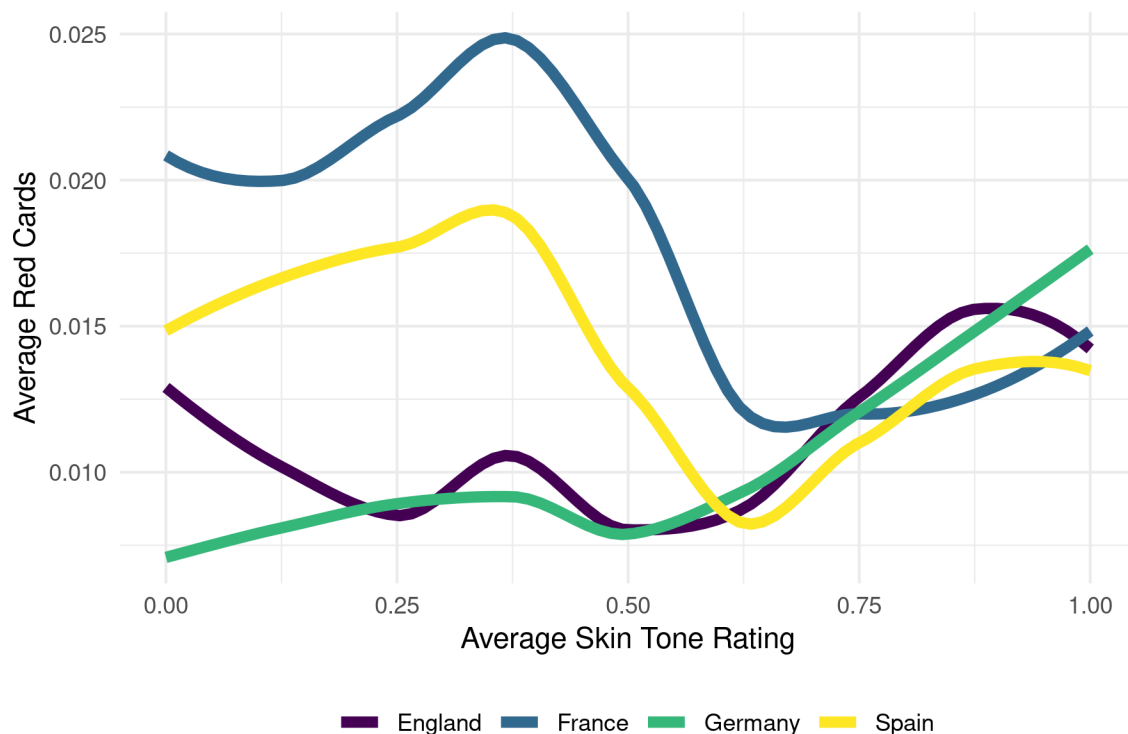
avg_red_cards_league <- clean_data %>%
  group_by(leagueCountry, mean_rating) %>%
  summarise(average_reds = mean(redCards))

ggplot(avg_red_cards_league, aes(x = mean_rating, y = average_reds, group = leagueCountry,
  geom_smooth(se = FALSE, size = 2) +
  scale_color_viridis(discrete = TRUE, option = "D") +
  theme_minimal() +
  labs(x = "Average Skin Tone Rating",
    y = "Average Red Cards",
    title = "Average Red Cards by Skin Tone Rating Across Leagues",
    subtitle = "Germany and England show an upward trend, while Spain and France tend to",
  theme(legend.title = element_blank(),
    legend.position = "bottom",
    plot.title = element_text(hjust = 0),
    plot.subtitle = element_text(hjust = 0, size = 10),
    plot.caption = element_text(hjust = 0.5),
    plot.title.position = "plot")

```

## Average Red Cards by Skin Tone Rating Across Leagues

Germany and England show an upward trend, while Spain and France tend to decrease



```

eng_ger_ref_data <- clean_data %>%
  filter(leagueCountry %in% c("England", "Germany"), redCards > 0) %>%
  group_by(refNum) %>%
  summarise(avg_meanIAT = mean(meanIAT, na.rm = TRUE)) %>%
  ungroup()

avg_meanIAT_eng_ger <- mean(eng_ger_ref_data$avg_meanIAT)

other_ref_data <- clean_data %>%
  filter(!leagueCountry %in% c("England", "Germany"), redCards > 0) %>%
  group_by(refNum) %>%
  summarise(avg_meanIAT = mean(meanIAT, na.rm = TRUE)) %>%
  ungroup()

avg_meanIAT_other <- mean(other_ref_data$avg_meanIAT)

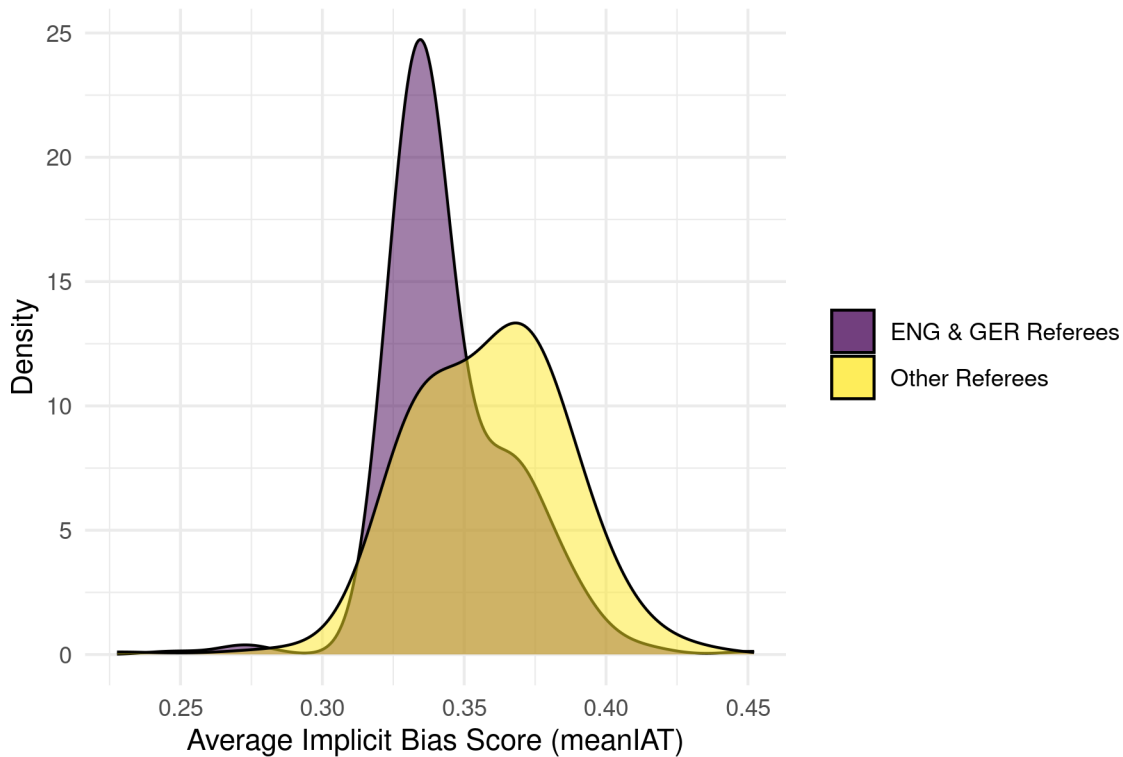
avg_data <- data.frame(
  League = c("England & Germany", "Other Leagues"),
  avg_meanIAT = c(avg_meanIAT_eng_ger, avg_meanIAT_other)
)

ggplot() +
  geom_density(data = eng_ger_ref_data, aes(x = avg_meanIAT, fill = "ENG & GER Referees"),
  geom_density(data = other_ref_data, aes(x = avg_meanIAT, fill = "Other Referees"), alpha
  scale_fill_viridis(discrete = TRUE) +
  labs(title = "Density of Referee Implicit Bias Scores (meanIAT)",
       subtitle = "Referees issuing red cards in GER and ENG do not exhibit bias",
       x = "Average Implicit Bias Score (meanIAT)",
       y = "Density") +
  theme_minimal() +
  theme(legend.title = element_blank(),
       plot.subtitle = element_text(size = 10, hjust = 0),
       plot.title.position = "plot")

```

## Density of Referee Implicit Bias Scores (meanIAT)

Referees issuing red cards in GER and ENG do not exhibit bias



- First Visual (Bar Plot: Distribution of Red Cards Across Skin Tone Ratings) This bar plot displays the average number of red cards given to players across different average skin tone ratings, ranging from 0 (lightest) to 1 (darkest).
- Second Visual (Line Plot: Average Red Cards by Skin Tone Rating Across Leagues) This line plot compares the trends in average red cards given within different European football leagues relative to players' average skin tone ratings.
- Third Visual (Density Plot: Density of Referee Implicit Bias Scores (meanIAT)) This density plot compares the implicit bias scores (meanIAT) of referees who gave red cards in German and English leagues against other referees.

## Second approach

```
median_games <- median(clean_data$games)
high_game_players <- clean_data %>%
  filter(games > median_games)
```

```

high_game_players <- high_game_players %>%
  mutate(red_card_rate = redCards / games)

avg_red_card_rate <- high_game_players %>%
  group_by(rater1) %>%
  summarise(average_red_card_rate = mean(red_card_rate))

plot1 <- ggplot(avg_red_card_rate, aes(x = rater1, y = average_red_card_rate, group = 1))
  geom_smooth(method = "loess", se = FALSE, color = "#440154", size = 1.2) +
  theme_minimal() +
  theme(plot.title.position = "plot") +
  labs(x = "Skin Tone Rating by Rater 1",
       y = "Average Red Card Rate")

high_game_players$skin_tone_category <- ifelse(high_game_players$mean_rating <= 0.2, "light",
                                              ifelse(high_game_players$mean_rating >= 0.8, "dark", "medium"))

high_game_players <- na.omit(high_game_players)

avg_red_card_rate_by_skin_tone <- high_game_players %>%
  group_by(skin_tone_category) %>%
  summarise(average_red_card_rate = mean(red_card_rate, na.rm = TRUE))

avg_red_card_rate_by_skin_tone$skin_tone_category <- factor(avg_red_card_rate_by_skin_tone$skin_tone_category,
                                                            levels = c("light skin", "dark skin", "medium skin"))

plot2 <- ggplot(avg_red_card_rate_by_skin_tone, aes(x = skin_tone_category, y = average_red_card_rate))
  geom_bar(stat = "identity", color = "black") +
  scale_fill_viridis(discrete = TRUE, direction = -1) +
  theme_minimal() +
  labs(x = "Skin Tone Category",
       y = "Average Red Card Rate") +
  theme(legend.position = "none",
        plot.title.position = "plot",
        plot.subtitle = element_text(size = 10))

position_red_cards <- clean_data %>%
  group_by(position) %>%
  summarise(total_red_cards = sum(redCards)) %>%

```

```

ungroup()

position_red_cards <- position_red_cards %>%
  arrange(desc(total_red_cards))

plot3 <- ggplot(position_red_cards, aes(x = reorder(position, total_red_cards), y = total_
  geom_bar(stat = "identity", color = "black", fill = "grey80") +
  coord_flip() +
  theme_minimal() +
  labs(x = "Position", y = "Total Red Cards") +
  theme(plot.title.position = "plot",
    axis.title.y = element_blank(),
    plot.subtitle = element_text(size = 10))

filtered_data <- clean_data %>%
  filter(leagueCountry %in% c("Germany", "England"),
    position %in% c("Center Back", "Defensive Midfielder", "Center Forward"))

position_skin_tones <- filtered_data %>%
  group_by(position) %>%
  summarise(average_skin_tone = mean(mean_rating, na.rm = TRUE),
    skin_tone_sd = sd(mean_rating, na.rm = TRUE)) %>%
  ungroup()

plot4 <- ggplot(position_skin_tones, aes(x = position, y = average_skin_tone, fill = facto
  geom_bar(stat = "identity", color = "black") +
  scale_fill_manual(values = c("Center Back" = "#fde725", "Defensive Midfielder" = "#5ec96
  theme_minimal() +
  labs(x = "Position", y = "Average Skin Tone Rating") +
  theme(legend.position = "none",
    legend.text = element_text(size = 8),
    plot.title.position = "plot",
    axis.title.y = element_blank(),
    axis.text.x = element_text(angle = 45, vjust = 1, hjust = 0.8, size = 8),
    plot.subtitle = element_text(size = 10))

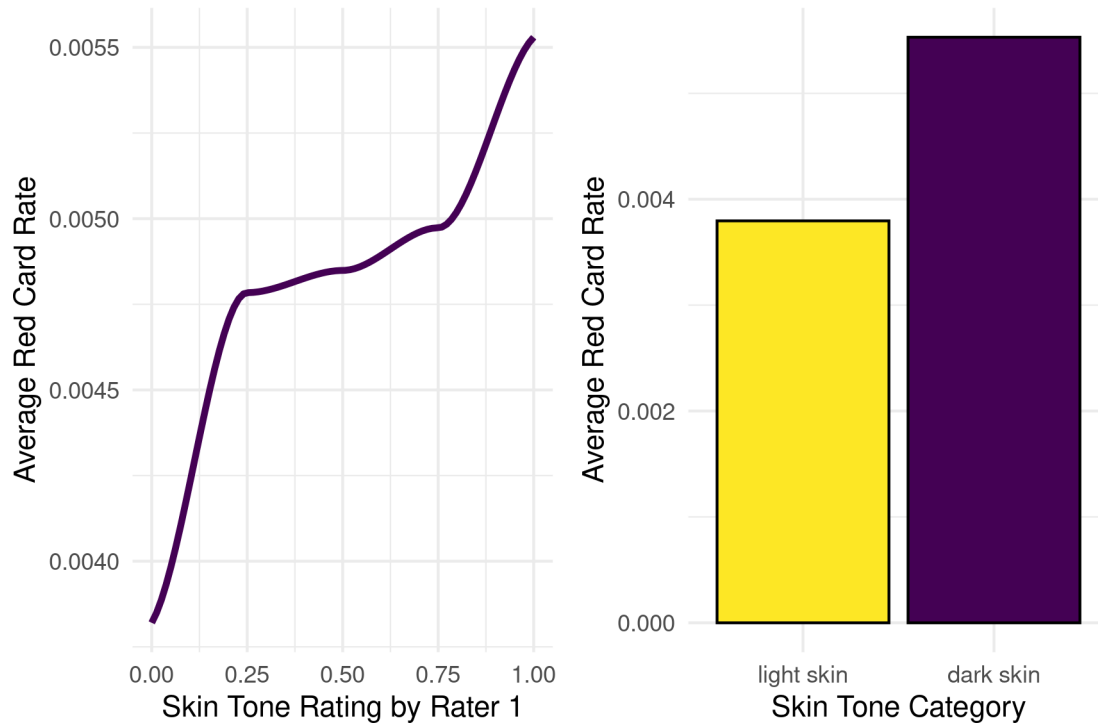
combined_plot1 <- plot1 + plot2
combined_plot1 + plot_annotation(
  title = "Comparative Analysis of Red Card Rates and Skin Tone",
  subtitle = "Evaluating Trends and Categorical Differences Among High-Game Players"

```

)

## Comparative Analysis of Red Card Rates and Skin Tone

### Evaluating Trends and Categorical Differences Among High-Game Players

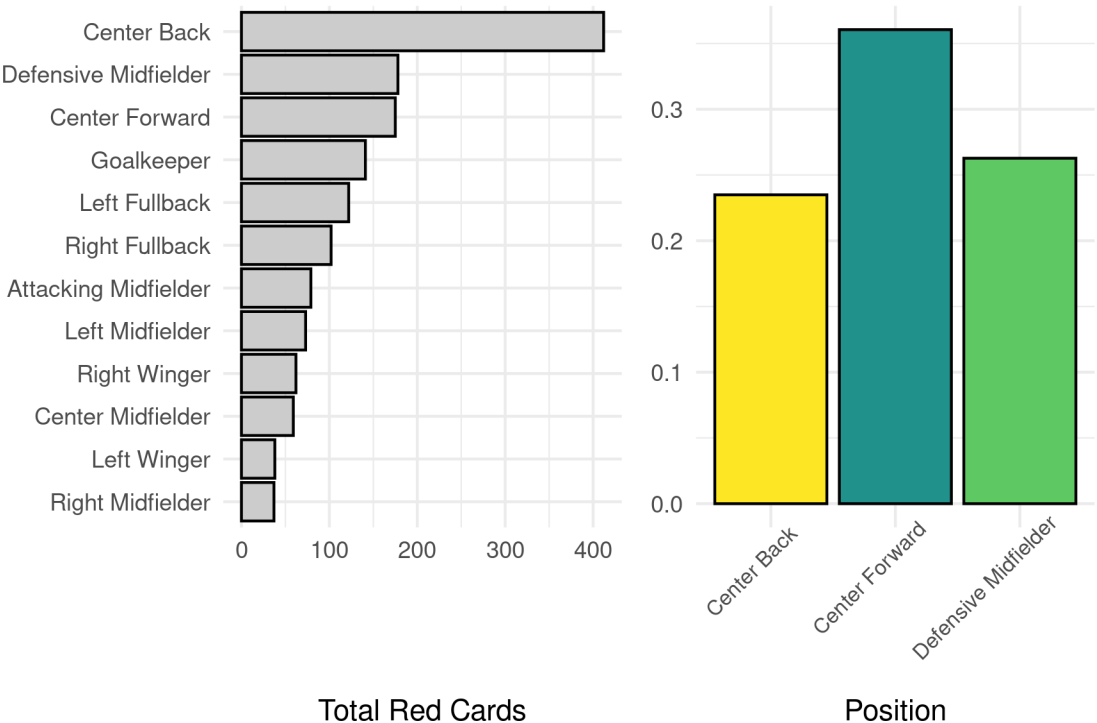


```
combined_plot2 <- plot3 + plot4
combined_plot2 + plot_annotation(
  title = "Player Position Impact on Red Cards and Correlation with Skin Tone",
  subtitle = "Predominance of Red Cards Among Roles and Their Skin Tone Ratings in Germany"
) &
theme(
  plot.subtitle = element_text(size = 9))
```



# Player Position Impact on Red Cards and Correlation with Skin Tone

Predominance of Red Cards Among Roles and Their Skin Tone Ratings in Germany and England



- First Visual (Line and Bar Plot: Red Card Rate vs. Skin Tone Rating)

This visual combines a line graph showing the average red card rate across different skin tone ratings and a bar chart comparing the average red card rate between light and dark skin categories among players with high game appearances.

- Second Visual (Bar Plot: Total Red Cards by Position & Skin Tone in Specific Leagues)

This visual showcases a bar chart detailing the total red cards received by player position, paired with a bar chart showing the average skin tone for positions with the most red cards in the German and English leagues.

## Reference

Open access | Research article | First published online August 23, 2018

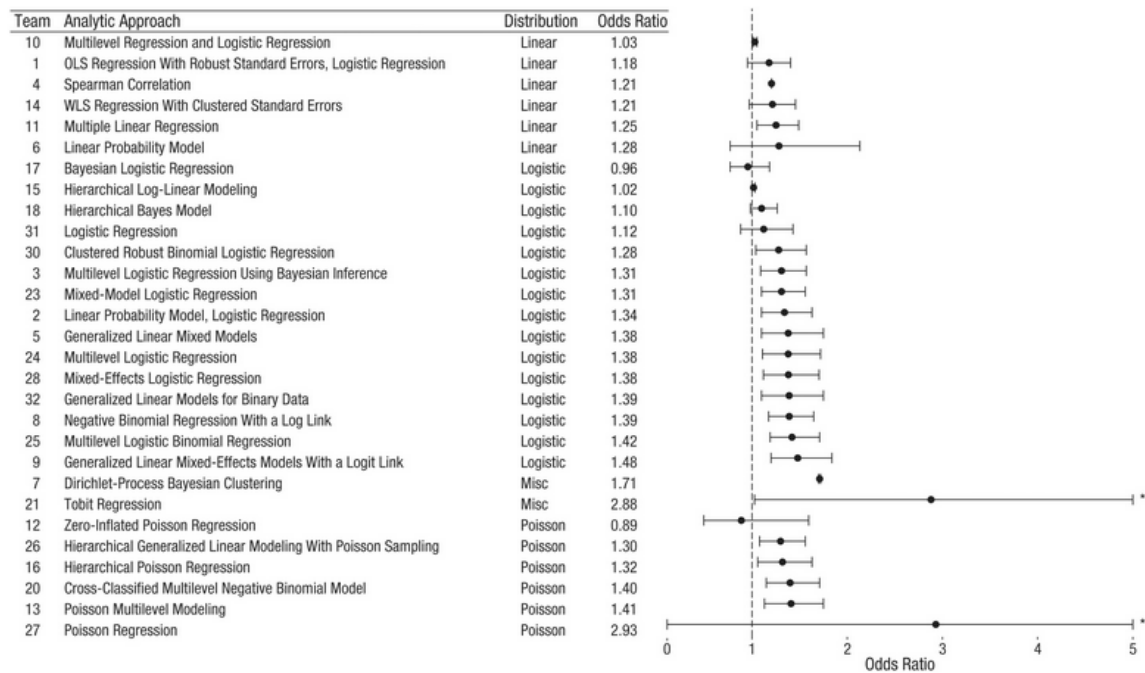
### Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results

R. Silberzahn, E. L. Uhlmann, and B. A. Nosek [View all authors and affiliations](#)

[All Articles](#) | <https://doi.org/10.1177/2515245917747646> | [View correction](#)

#### Abstract

Twenty-nine teams involving 61 analysts used the same data set to address the same research question: whether soccer referees are more likely to give red cards to dark-skin-toned players than to light-skin-toned players. Analytic approaches varied widely across the teams, and the estimated effect sizes ranged from 0.89 to 2.93 (*Mdn* = 1.31) in odds-ratio units. Twenty teams (69%) found a statistically significant positive effect, and 9 teams (31%) did not observe a significant relationship. Overall, the 29 different analyses used 21 unique combinations of covariates. Neither analysts' prior beliefs about the effect of interest nor their level of expertise readily explained the variation in the outcomes of the analyses. Peer ratings of the quality of the analyses also did not account for the variability. These findings suggest that significant variation in the results of analyses of complex data may be difficult to avoid, even by experts with honest intentions. Crowdsourcing data analysis, a strategy in which numerous research teams are recruited to simultaneously investigate the same research question, makes transparent how defensible, yet subjective, analytic choices influence research results.



- <https://journals.sagepub.com/doi/10.1177/2515245917747646>