31_Data_Visualization

Group 20

2025-11-07

Processing

```
# 1. Load datasets
regular <- read_csv("Regular Season.csv")
```

```
## New names:
## Rows: 570 Columns: 15
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (2): Player, Team dbl (13): ...1, Age, GP, W, L, Min, PTS, FGM, FGA, FG%, 3PM/,
## 3PA, 3P%
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
playoffs <- read_csv("Playoffs.csv")
```

```
## New names:
## Rows: 219 Columns: 15
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (2): Player, Team dbl (13): ...1, Age, GP, W, L, Min, PTS, FGM, FGA, FG%, 3PM/,
## 3PA, 3P%
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
accuracy_comparison <- read_csv("team_accuracy_comparison_R.csv")
```

```
## Rows: 30 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Team, Team_Status
## dbl (2): FG_percent_team, TP_percent_team
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# 2. Keep only relevant columns
regular_clean <- regular %>%
  select(Player, Team, GP, Min, PTS, FGM, FGA, `FG%`, `3PM/`, `3PA`, `3P%`)

playoffs_clean <- playoffs %>%
  select(Player, Team, GP, Min, PTS, FGM, FGA, `FG%`, `3PM/`, `3PA`, `3P%`)

# 3. Rename columns for consistency
names(regular_clean) <- c("Player", "Team", "GP", "Minutes", "Points",
                          "FGM", "FGA", "FG_percent", "TPM", "TPA", "TP_percent")

names(playoffs_clean) <- c("Player", "Team", "GP", "Minutes", "Points",
                           "FGM", "FGA", "FG_percent", "TPM", "TPA", "TP_percent")

# 4. Process: Create a paired dataset
# We use an inner_join to keep only players who appear in z` datasets
paired_stats <- inner_join(
  regular_clean,
  playoffs_clean,
  by = "Player",
  suffix = c("_reg", "_playoff")
)

# Glimpse the final paired data
glimpse(paired_stats)
```

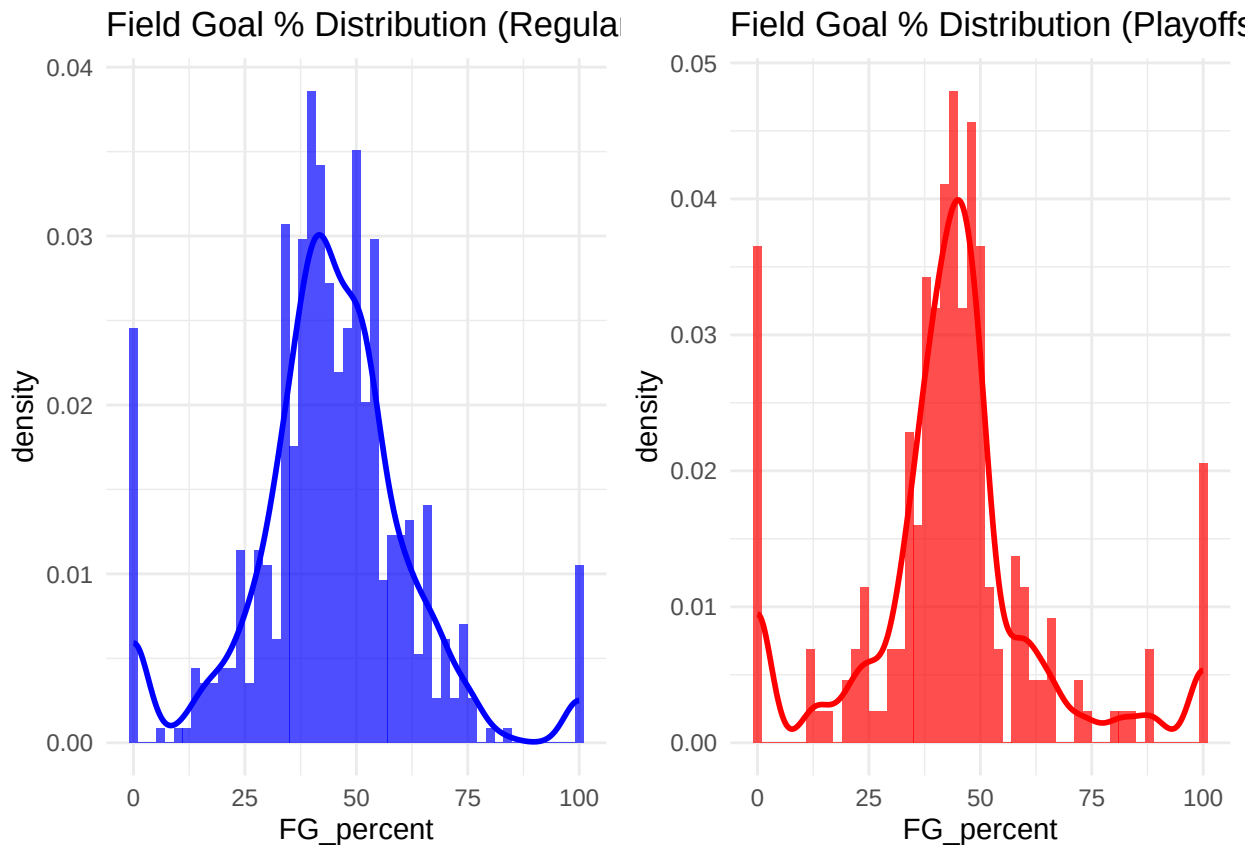
```
## Rows: 204
## Columns: 21
## $ Player             <chr> "Al Horford", "Baylor Scheierman", "Derrick White",~
## $ Team_reg           <chr> "BOS", "BOS", "BOS", "BOS", "BOS", "BOS", "BOS", "B~
## $ GP_reg             <dbl> 1, 5, 4, 5, 4, 4, 5, 3, 4, 5, 5, 4, 3, 4, 3, 4, 3, ~
## $ Minutes_reg        <dbl> 23.1, 58.2, 96.0, 65.6, 90.5, 99.0, 104.0, 48.0, 74~
## $ Points_reg         <dbl> 2, 15, 50, 35, 66, 69, 45, 20, 36, 29, 80, 42, 15, ~
## $ FGM_reg            <dbl> 1, 4, 17, 10, 24, 24, 15, 9, 17, 10, 27, 15, 6, 9, ~
## $ FGA_reg            <dbl> 6, 24, 45, 22, 52, 58, 33, 17, 23, 17, 56, 33, 15, ~
## $ FG_percent_reg     <dbl> 16.7, 16.7, 37.8, 45.5, 46.2, 41.4, 45.5, 52.9, 73.~
## $ TPM_reg            <dbl> 0, 3, 9, 6, 4, 14, 7, 2, 0, 0, 21, 11, 3, 6, 0, 3, ~
## $ TPA_reg            <dbl> 4, 17, 28, 14, 21, 36, 19, 8, 0, 1, 43, 24, 10, 11,~
## $ TP_percent_reg     <dbl> 0.0, 17.6, 32.1, 42.9, 19.0, 38.9, 36.8, 25.0, 0.0,~
## $ Team_playoff       <chr> "BOS", "BOS", "BOS", "BOS", "BOS", "BOS", "BOS", "B~
## $ GP_playoff         <dbl> 11, 4, 11, 4, 11, 8, 5, 8, 11, 4, 11, 8, 5, 1, 3, 6~
## $ Minutes_playoff    <dbl> 347.6, 22.3, 415.3, 13.2, 402.4, 321.6, 15.3, 264.0~
## $ Points_playoff     <dbl> 88, 8, 207, 4, 243, 225, 0, 76, 50, 10, 131, 28, 6,~
## $ FGM_playoff        <dbl> 34, 3, 68, 1, 86, 74, 0, 29, 21, 5, 45, 10, 2, 1, 1~
## $ FGA_playoff        <dbl> 72, 10, 147, 8, 195, 175, 1, 60, 29, 6, 99, 24, 4, ~
## $ FG_percent_playoff <dbl> 47.2, 30.0, 46.3, 12.5, 44.1, 42.3, 0.0, 48.3, 72.4~
## $ TPM_playoff        <dbl> 14, 2, 40, 1, 21, 29, 0, 9, 0, 0, 27, 7, 2, 0, 1, 0~
## $ TPA_playoff        <dbl> 35, 5, 104, 2, 63, 78, 1, 26, 0, 0, 67, 21, 3, 2, 4~
## $ TP_percent_playoff <dbl> 40.0, 40.0, 38.5, 50.0, 33.3, 37.2, 0.0, 34.6, 0.0,~
```

Distribution of Shooting Efficiency

```
plot_distribution_histograms(
  regular_df = regular_clean,
  playoff_df = playoffs_clean,
  metric_col = "FG_percent",
  plot_title_prefix = "Field Goal %",
  bin_width = 2
)
```

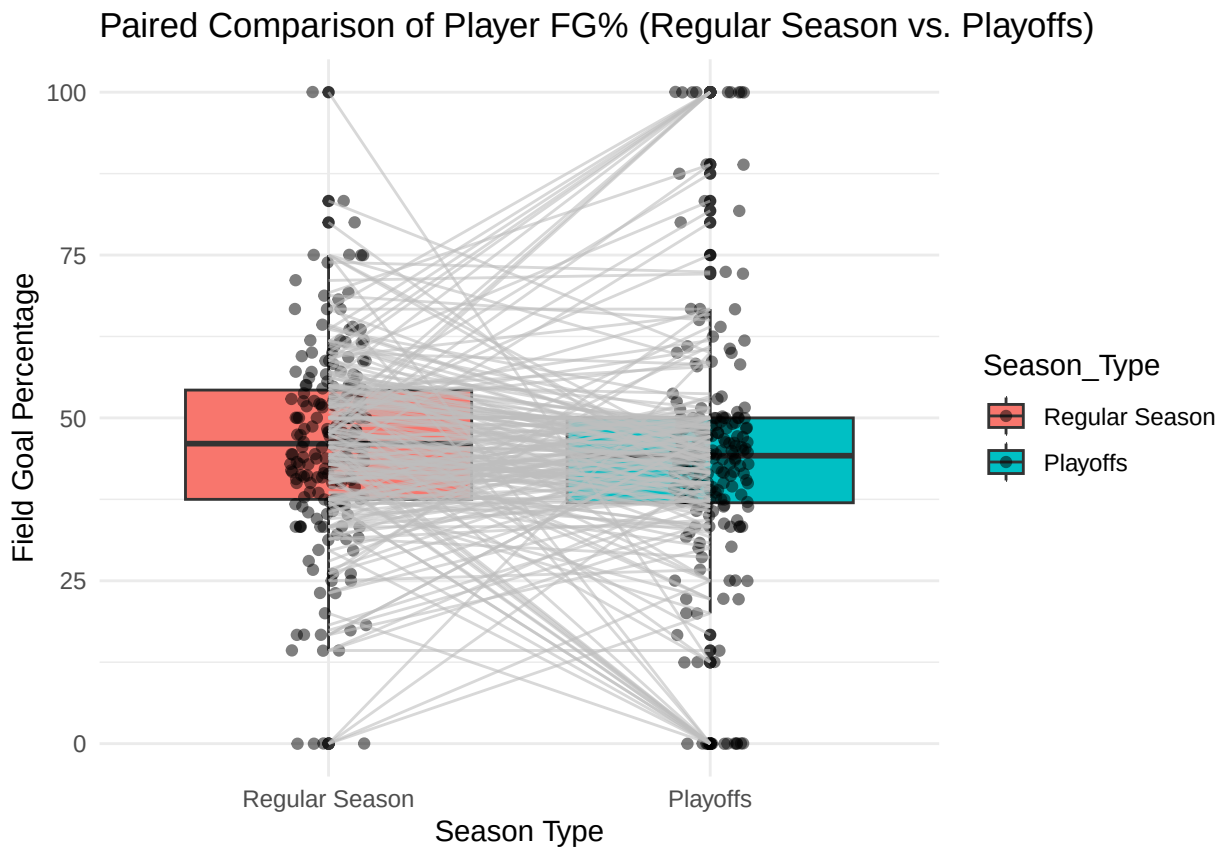
```
## Warning: ‘aes_string()‘ was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with ‘aes()‘.
## i See also ‘vignette("ggplot2-in-packages")‘ for more information.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
## 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.
```



Paired Comparison of Performance

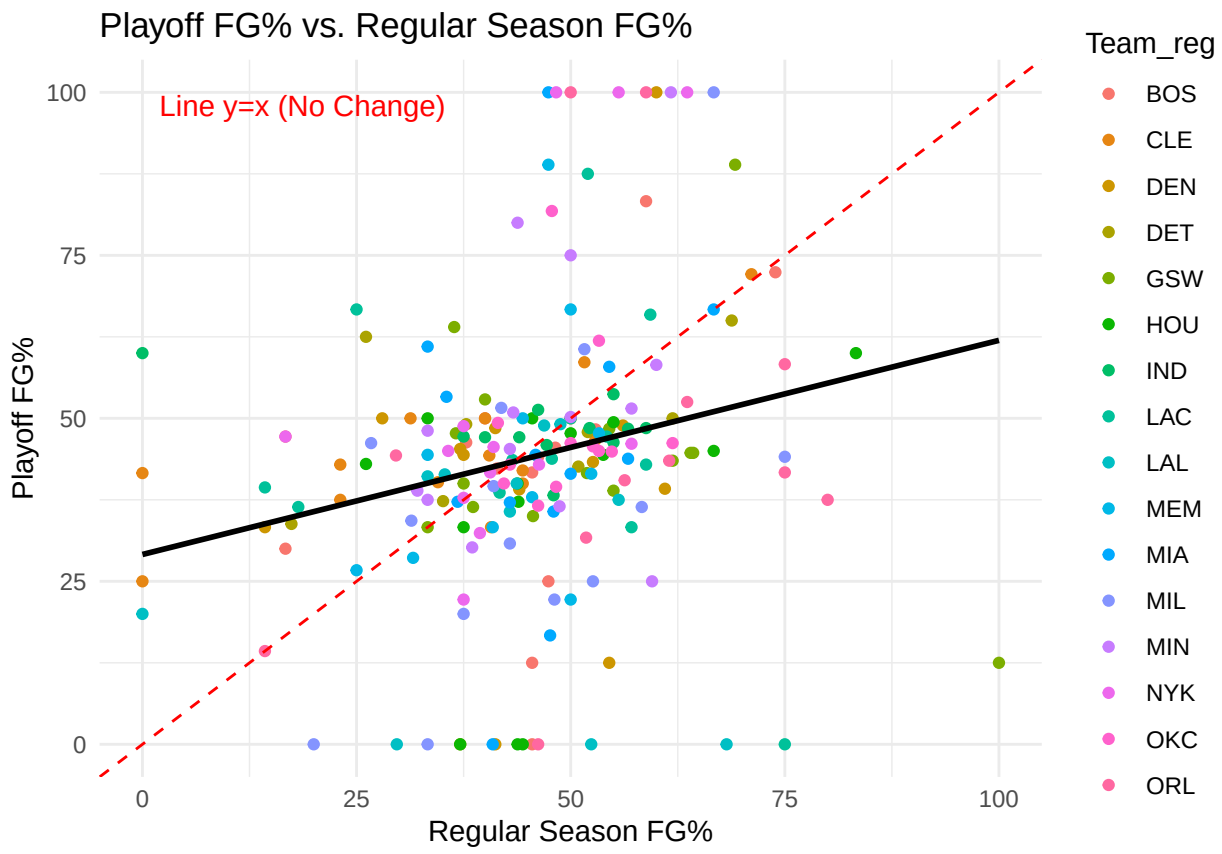
```
plot_paired_boxplot(
  paired_df = paired_stats,
  col_reg = "FG_percent_reg",
  col_playoff = "FG_percent_playoff",
  plot_title = "Paired Comparison of Player FG% (Regular Season vs. Playoffs)",
  y_label = "Field Goal Percentage"
)
```



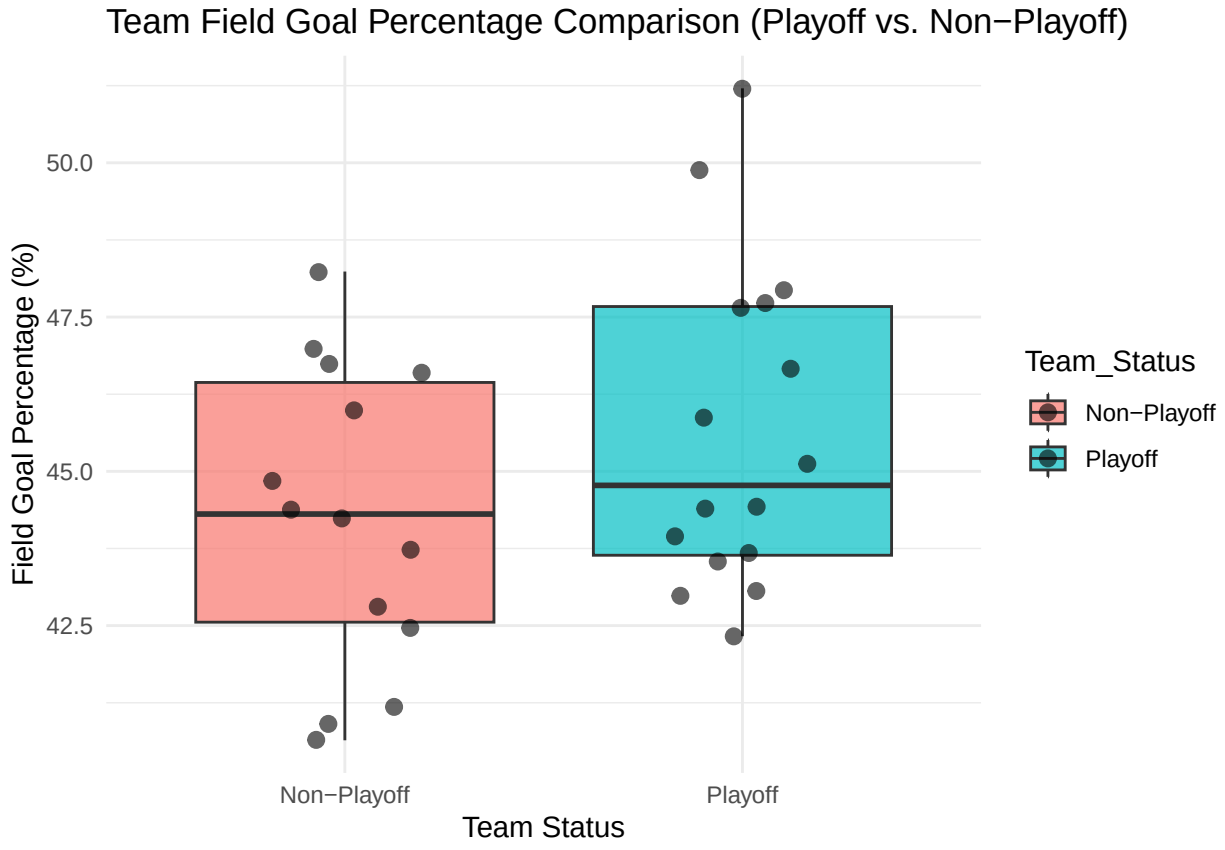
Correlation of Performance

```
plot_correlation_scatter(
  paired_df = paired_stats,
  col_reg = "FG_percent_reg",
  col_playoff = "FG_percent_playoff",
  plot_title = "Playoff FG% vs. Regular Season FG%",
  x_label = "Regular Season FG%",
  y_label = "Playoff FG%"
)
```

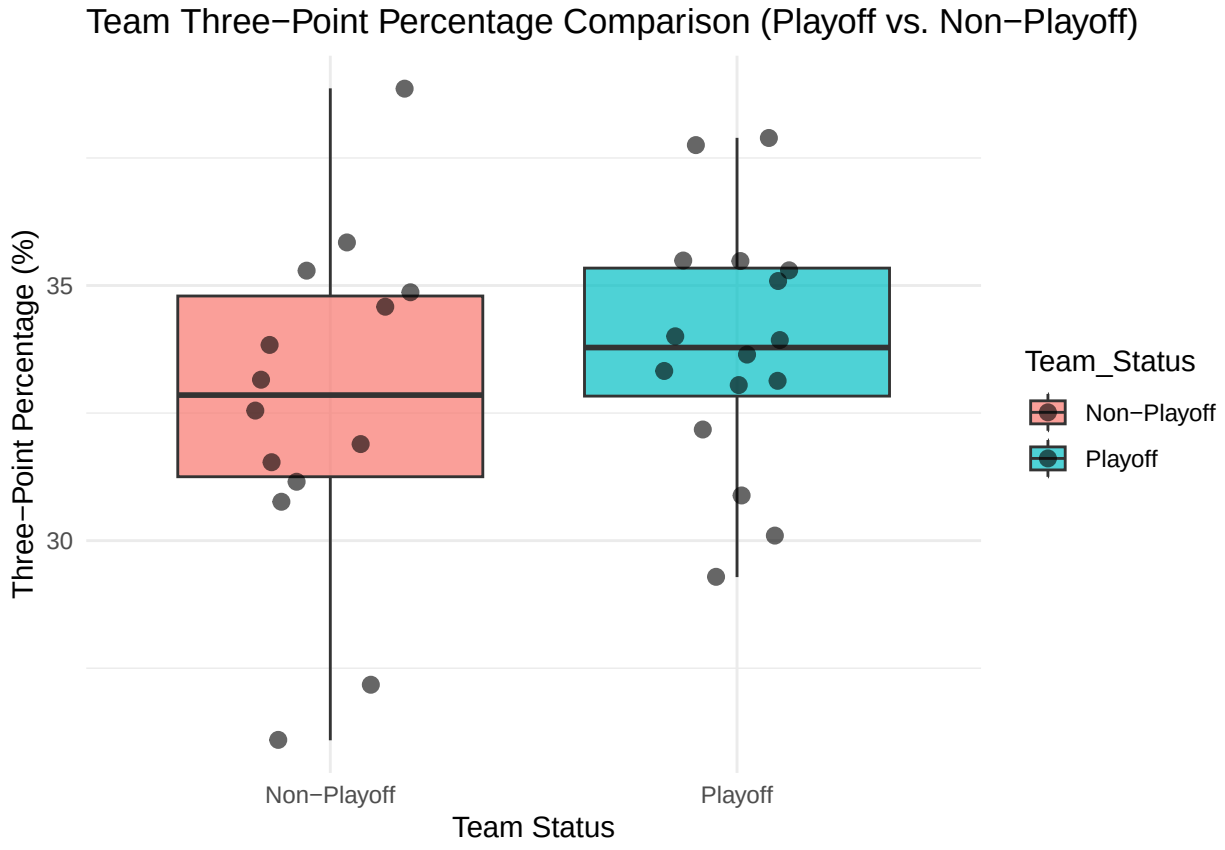
```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
# Box Plot for Field Goal Percentage by Team Status
# Compares the distribution of FG% between Playoff and Non-Playoff teams.
plot_fg_boxplot <- ggplot(accuracy_comparison,
                          aes(x = Team_Status, y = FG_percent_team, fill = Team_Status)) +
  geom_boxplot(alpha = 0.7) +
  # Add individual points (jitter) for visual clarity
  geom_jitter(color = "black", size = 2.5, width = 0.2, alpha = 0.6) +
  labs(
    title = "Team Field Goal Percentage Comparison (Playoff vs. Non-Playoff)",
    x = "Team Status",
    y = "Field Goal Percentage (%)"
  ) +
  theme_minimal() +
  # Use custom colors for distinct categories
  scale_fill_manual(values = c("Playoff" = "#00BFC4", "Non-Playoff" = "#F8766D"))

print(plot_fg_boxplot)
```

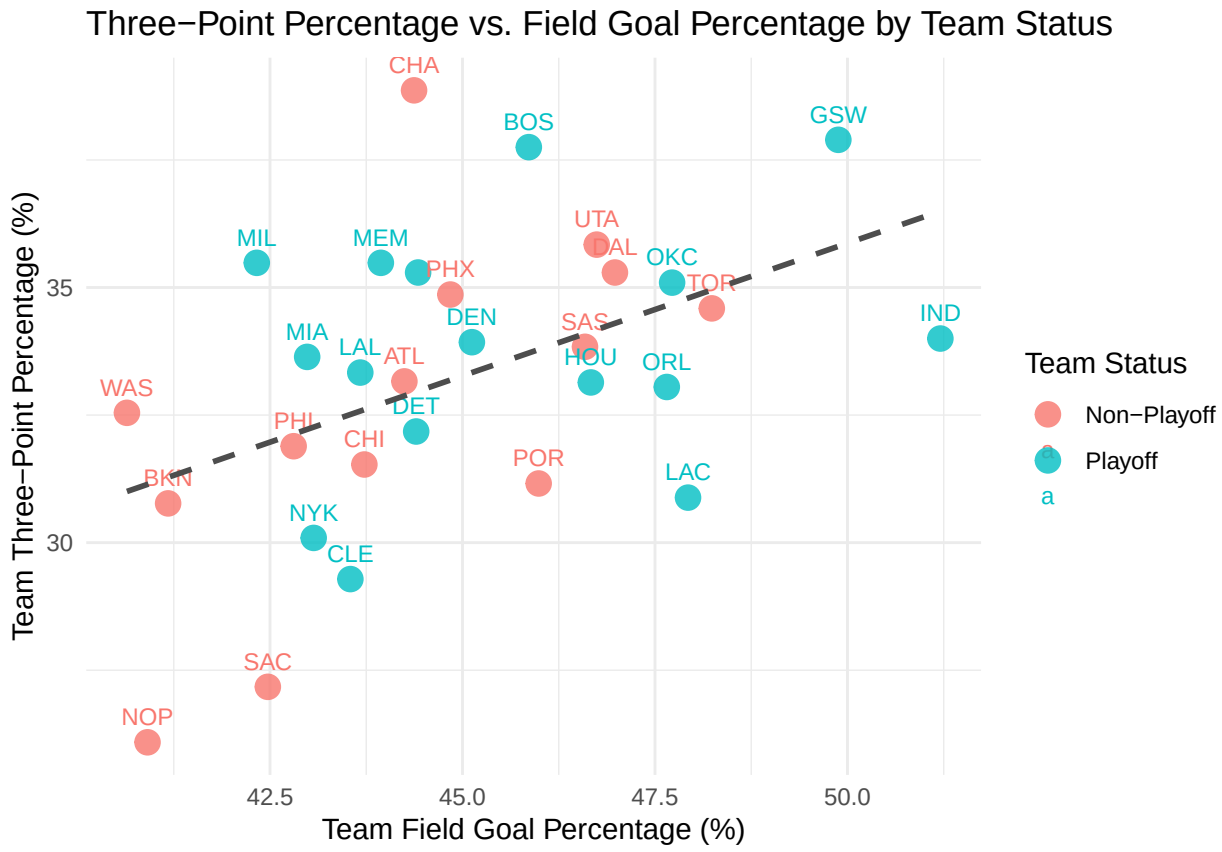



```r
# Box Plot for Three-Point Percentage by Team Status
# Compares the distribution of TP% between Playoff and Non-Playoff teams.
plot_tp_boxplot <- ggplot(accuracy_comparison,
                          aes(x = Team_Status, y = TP_percent_team, fill = Team_Status)) +
  geom_boxplot(alpha = 0.7) +
  # Add individual points (jitter) for visual clarity
  geom_jitter(color = "black", size = 2.5, width = 0.2, alpha = 0.6) +
  labs(
    title = "Team Three-Point Percentage Comparison (Playoff vs. Non-Playoff)",
    x = "Team Status",
    y = "Three-Point Percentage (%)"
  ) +
  theme_minimal() +
  scale_fill_manual(values = c("Playoff" = "#00BFC4", "Non-Playoff" = "#F8766D"))

print(plot_tp_boxplot)
```




```
# Scatter Plot of TP% vs. FG% with Team Labels
# Shows the correlation between the two metrics, colored by team status.
plot_scatter <- ggplot(accuracy_comparison,
                       aes(x = FG_percent_team, y = TP_percent_team, color = Team_Status)) +
  geom_point(size = 4, alpha = 0.8) +
  # Add team names as labels
  geom_text(aes(label = Team), vjust = -1, size = 3, check_overlap = TRUE) +
  # Add a linear regression line to show correlation
  geom_smooth(method = "lm", se = FALSE, color = "gray30", linetype = "dashed") +
  labs(
    title = "Three-Point Percentage vs. Field Goal Percentage by Team Status",
    x = "Team Field Goal Percentage (%)",
    y = "Team Three-Point Percentage (%)",
    color = "Team Status"
  ) +
  theme_minimal()

print(plot_scatter)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


Three−Point Percentage vs. Field Goal Percentage by Team Status
Team Three−Point Percentage (%)
35
30
42.5
45.0
47.5
50.0
Team Field Goal Percentage (%)
CHA
BOS
GSW
MIL
MEM
UTA
DAL
OKC
PHX
TOR
DEN
IND
MIA
SAS
LAL
ATL
HOU
ORL
WAS
DET
PHI
CHI
POR
LAC
BKN
NYK
CLE
SAC
NOP
Team Status
Non−Playoff
Playoff