31_Data_Visualization

Group 20

2025-11-07

Processing

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

```
## New names:
## Rows: 28 Columns: 14
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (2): Player, Team dbl (12): ...1, 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: 28 Columns: 14
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (2): Player, Team dbl (12): ...1, 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`
```

```
# 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 BOTH datasets
paired_stats <- inner_join(
  regular_clean,
  playoffs_clean,
  by = "Player",
  suffix = c("_reg", "_playoff")
)

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

```
## Rows: 28
## Columns: 21
## $ Player             <chr> "Shai Gilgeous-Alexander", "Jalen Williams", "Chet ~
## $ Team_reg           <chr> "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "O~
## $ GP_reg             <dbl> 76, 69, 32, 54, 57, 71, 76, 68, 74, 36, 47, 69, 54,~
## $ Minutes_reg        <dbl> 34.2, 32.4, 27.4, 19.3, 27.9, 29.2, 22.9, 27.6, 21.~
## $ Points_reg         <dbl> 32.7, 21.6, 15.0, 7.1, 11.2, 10.1, 12.0, 8.4, 10.2,~
## $ FGM_reg            <dbl> 11.3, 8.2, 5.2, 2.6, 4.9, 3.6, 4.7, 3.4, 3.5, 2.5, ~
## $ FGA_reg            <dbl> 21.8, 16.9, 10.7, 5.8, 8.4, 8.4, 9.6, 7.2, 7.9, 5.1~
## $ FG_percent_reg     <dbl> 51.9, 48.4, 49.0, 44.6, 58.1, 43.5, 48.8, 47.4, 44.~
## $ TPM_reg            <dbl> 2.1, 1.8, 1.4, 1.1, 0.0, 2.4, 1.7, 1.1, 2.6, 0.6, 1~
## $ TPA_reg            <dbl> 5.7, 4.9, 3.6, 3.1, 0.3, 5.8, 4.5, 3.1, 6.3, 1.7, 3~
## $ TP_percent_reg     <dbl> 37.5, 36.5, 37.9, 35.3, 0.0, 41.2, 38.3, 35.6, 41.2~
## $ Team_playoff       <chr> "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "OKC", "O~
## $ GP_playoff         <dbl> 23, 23, 23, 23, 23, 23, 22, 23, 21, 12, 17, 16, 10,~
## $ Minutes_playoff    <dbl> 37.0, 34.6, 29.8, 24.4, 22.4, 28.9, 13.8, 22.4, 10.~
## $ Points_playoff     <dbl> 29.9, 21.4, 15.2, 9.2, 8.1, 7.9, 6.0, 5.6, 5.1, 3.4~
## $ FGM_playoff        <dbl> 10.1, 7.7, 5.3, 3.1, 3.6, 2.6, 2.1, 2.2, 1.7, 1.3, ~
## $ FGA_playoff        <dbl> 21.9, 17.2, 11.6, 7.0, 5.8, 7.1, 5.4, 5.2, 3.5, 2.9~
## $ FG_percent_playoff <dbl> 46.2, 44.9, 46.2, 45.0, 61.9, 36.6, 39.5, 42.9, 49.~
## $ TPM_playoff        <dbl> 1.4, 1.5, 1.2, 1.6, 0.0, 2.1, 1.1, 0.9, 1.1, 0.4, 0~
## $ TPA_playoff        <dbl> 4.9, 5.0, 4.0, 3.9, 0.0, 6.1, 3.1, 2.7, 2.7, 1.1, 1~
## $ TP_percent_playoff <dbl> 28.3, 30.4, 29.7, 41.1, 0.0, 34.3, 36.2, 32.3, 41.1~
```

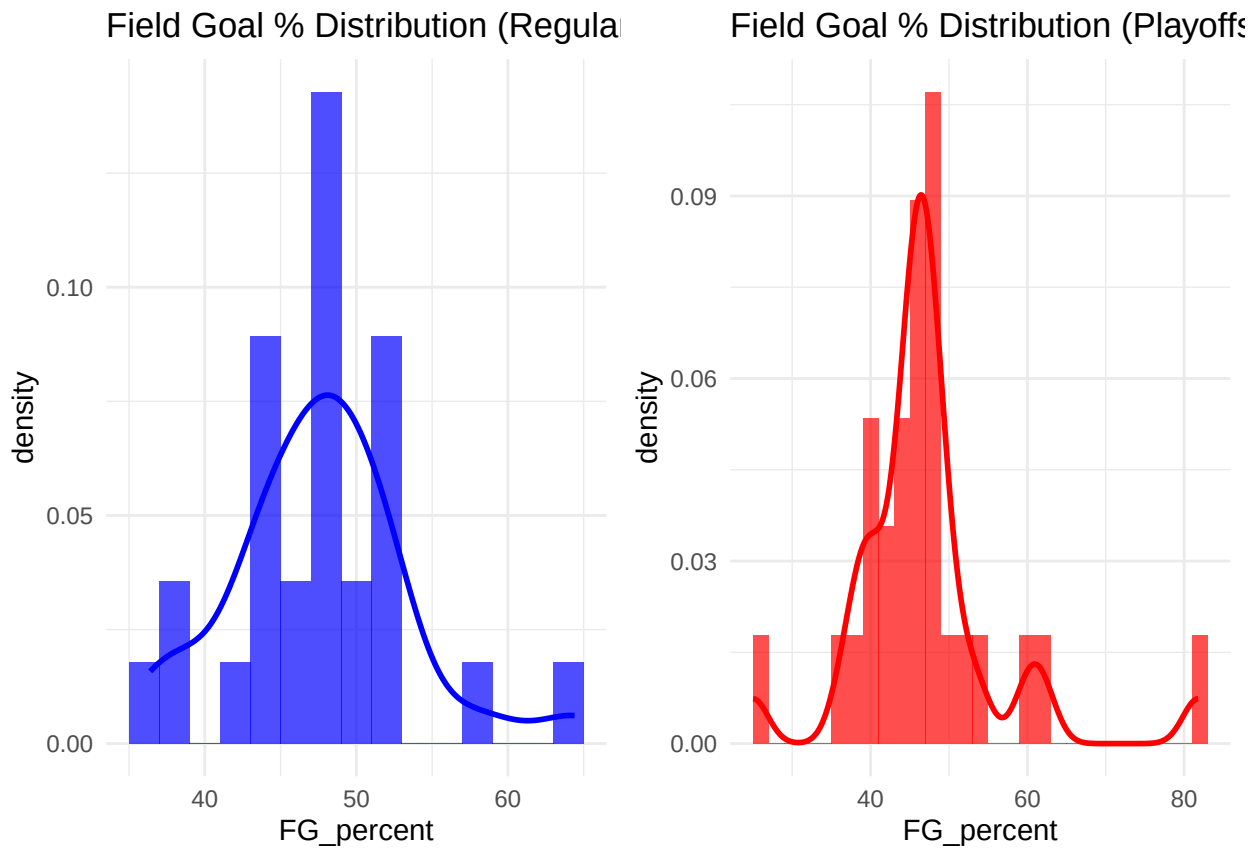
Distribution of Shooting Efficiency

```
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.
```

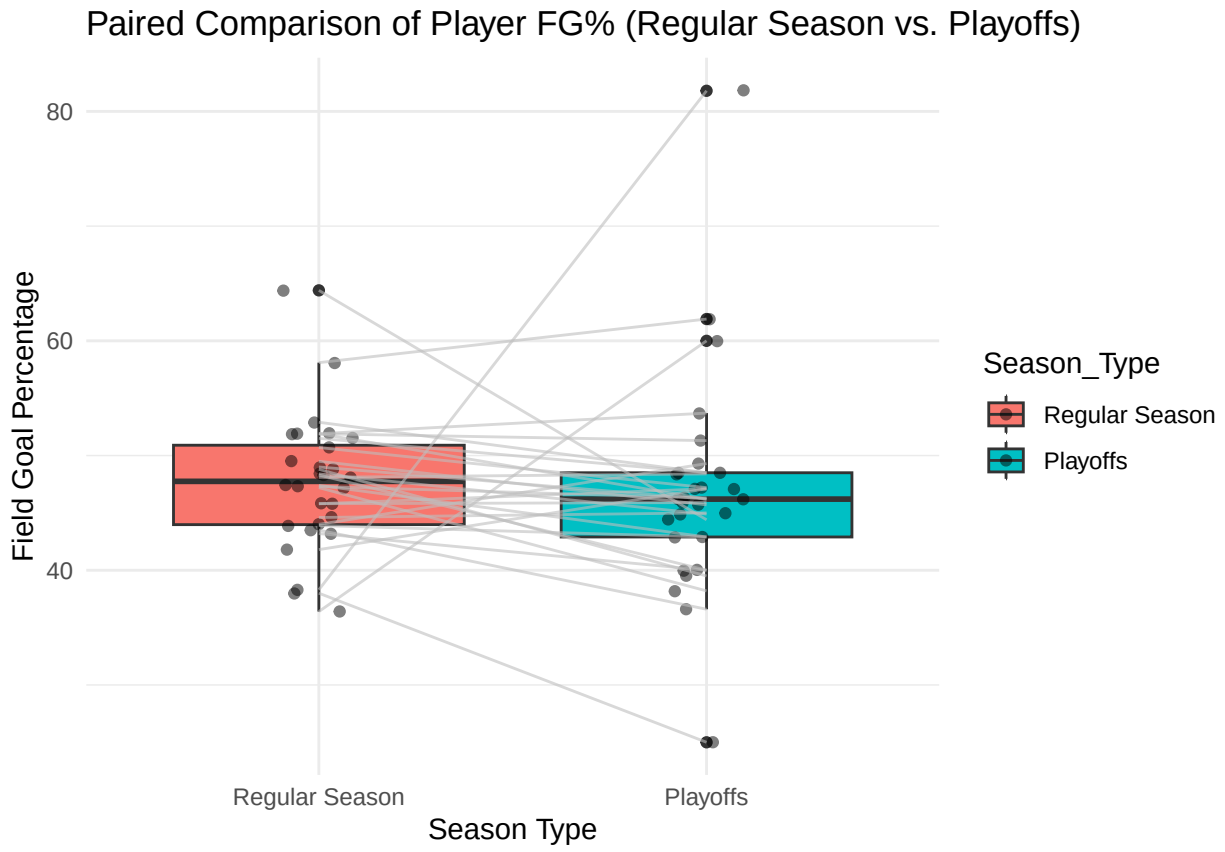
```
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.

## 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

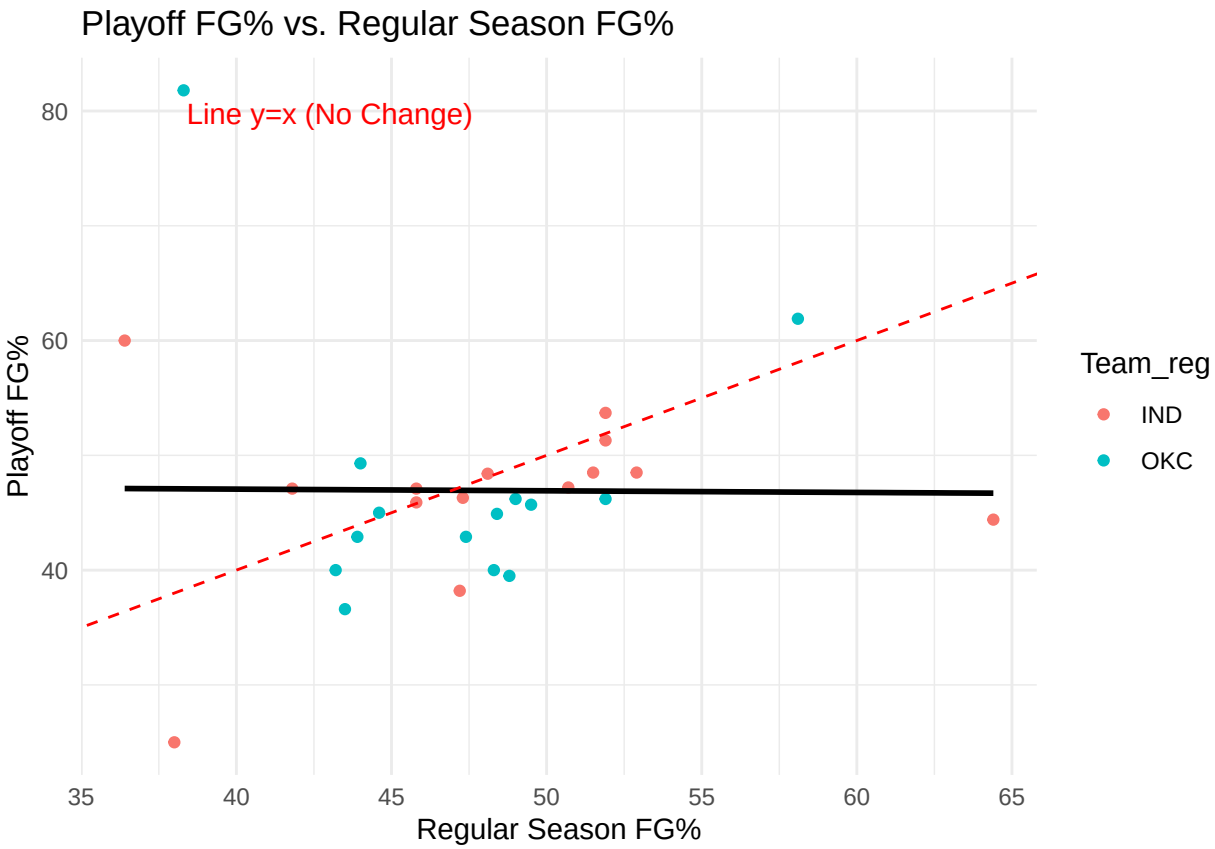
```
plot_paired_boxplot(
  paired_df = paired_stats,
  col_reg = "FG_percent_reg",
  col_playoff = "FG_percent_playoff",
  plot_title = "Paired Comparison of Player FG% (Regular Season vs. Playoffs)",
  y_label = "Field Goal Percentage"
)
```



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'
```


Playoff FG% vs. Regular Season FG%
Line y=x (No Change)
Playoff FG%
Regular Season FG%
80
60
40
35
40
45
50
55
60
65
Team_reg
IND
OKC