

SUMMER 2024

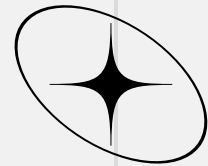
CS 544

ANALYSIS OF BOSTON CELTICS REGULAR SEASON SHOOTING

2022-23

KADER NDOYE





DATASETS USED

<https://www.kaggle.com/datasets/johnmimnaugh/boston-celtics-shooting-variables>

BOXSCORE TOTALS

Shows traditional box score stats

DRIBBLES SHOOTING

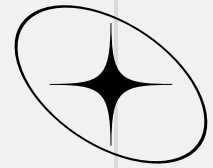
Shows how many dribbles players made before shooting the ball, along with shooting percentages

SHOT CLOCK SHOOTING

Shows shooting percent in different ranges of the shot clock.

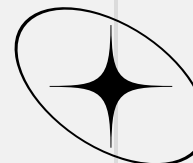
TOUCH TIME SHOOTING

Shows how long a player had possession of the ball before attempting a shot, along with shooting percentages

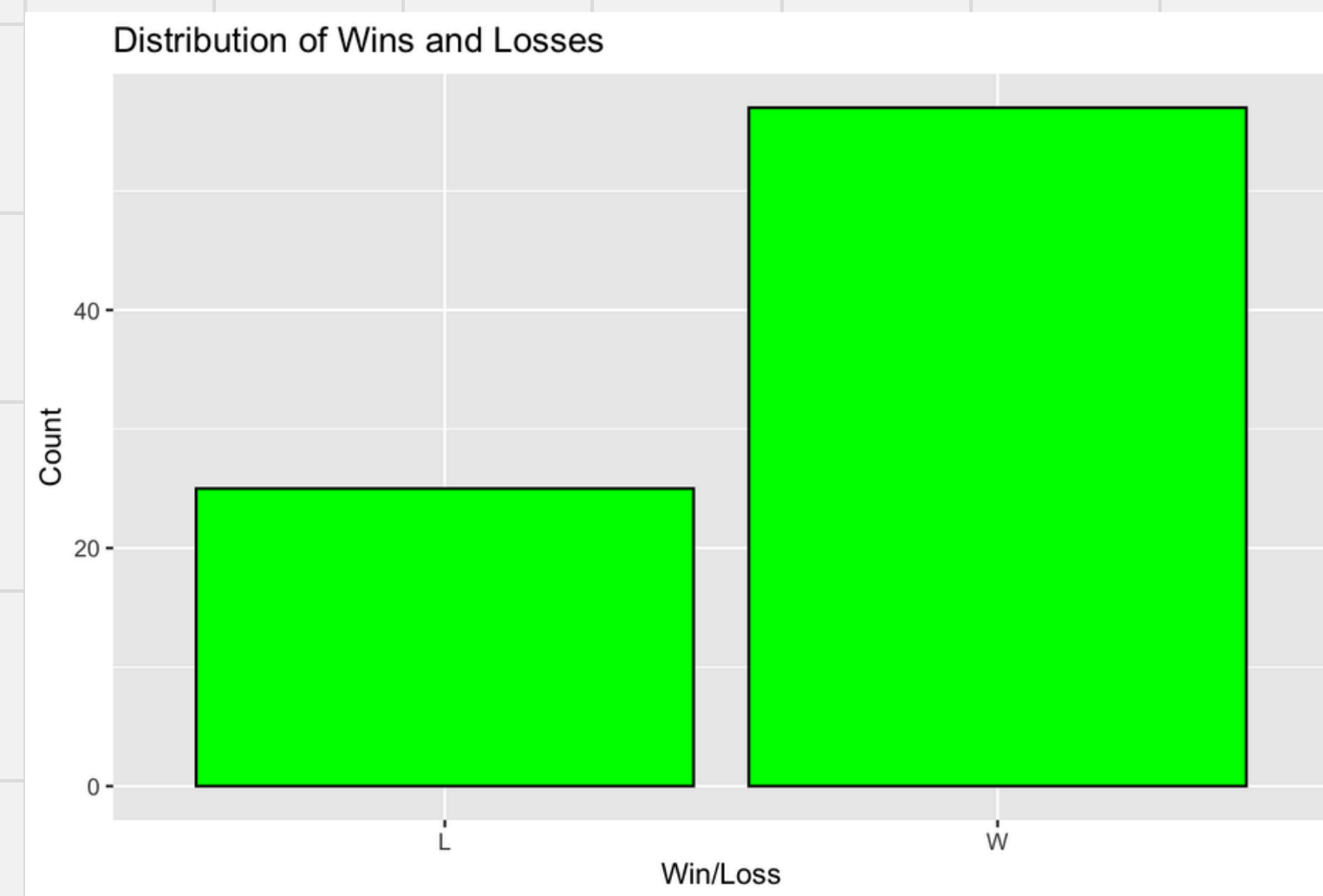
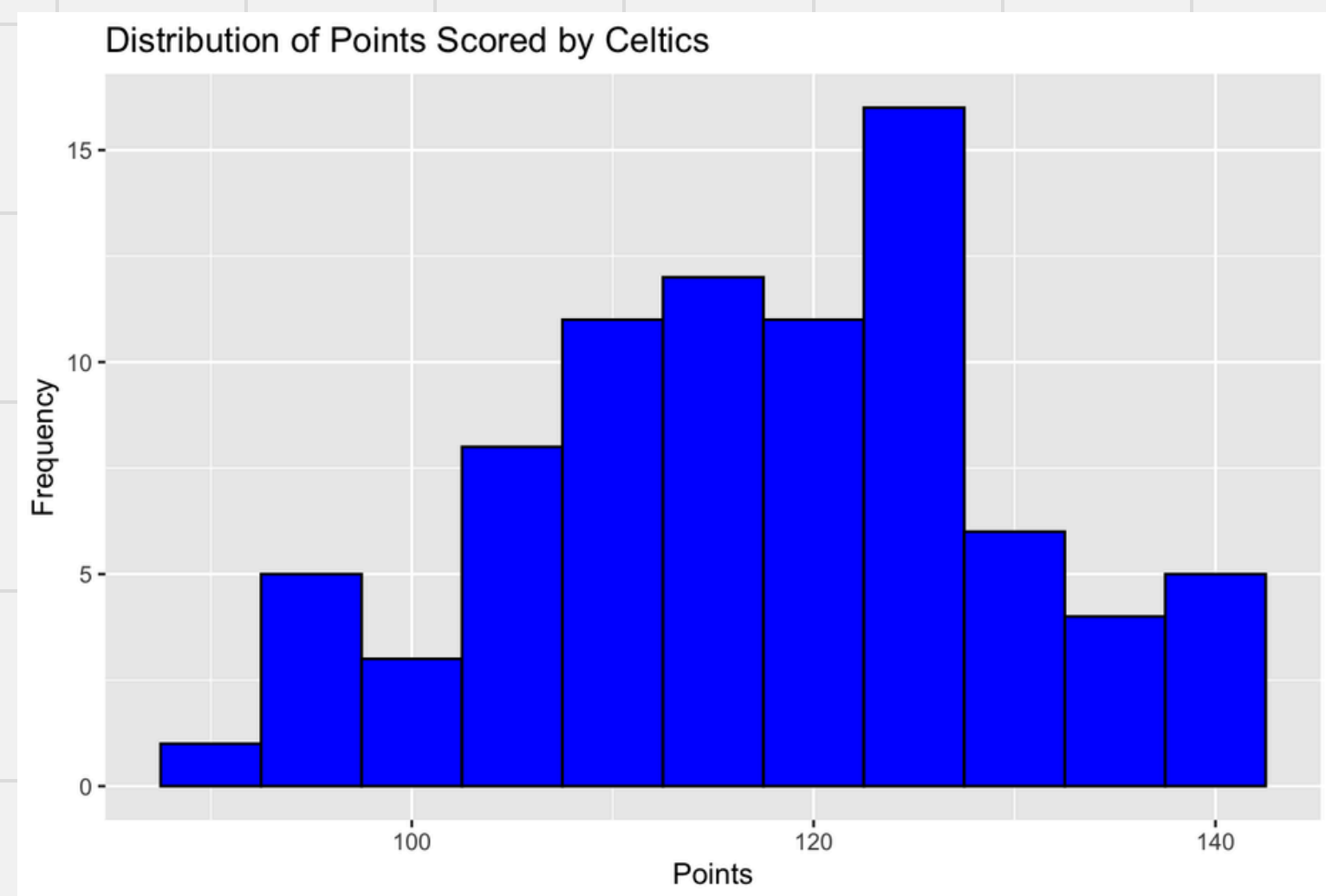


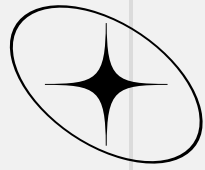
```
> # Check for missing values and clean datasets
> missing_values <- sum(is.na(boxscore_totals))
> cat("Number of missing values:", missing_values, "\n")
Number of missing values: 0
> # Check for missing values and clean datasets
> missing_values <- sum(is.na(boxscore_totals))
> cat("Number of missing values:", missing_values, "\n")
Number of missing values: 0
> missing_values <- sum(is.na(dribbles_shooting))
> cat("Number of missing values:", missing_values, "\n")
Number of missing values: 0
> missing_values <- sum(is.na(shot_clock_shooting))
> cat("Number of missing values:", missing_values, "\n")
Number of missing values: 0
> missing_values <- sum(is.na(touch_time_shooting))
> cat("Number of missing values:", missing_values, "\n")
Number of missing values: 0
```

DATA PREPARATION



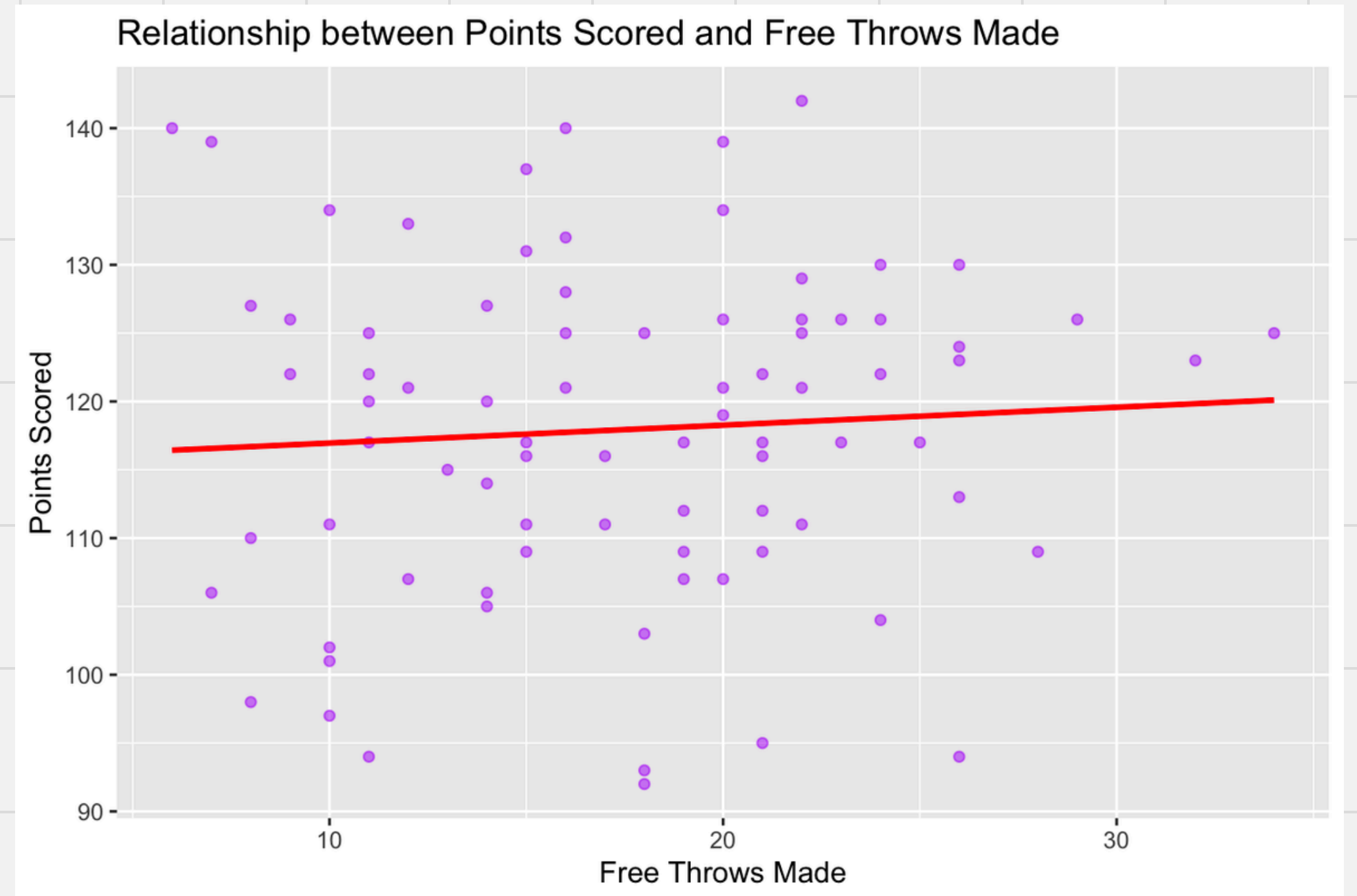
SUMMARY STATISTICS





RELATIONSHIP BETWEEN VARIABLES

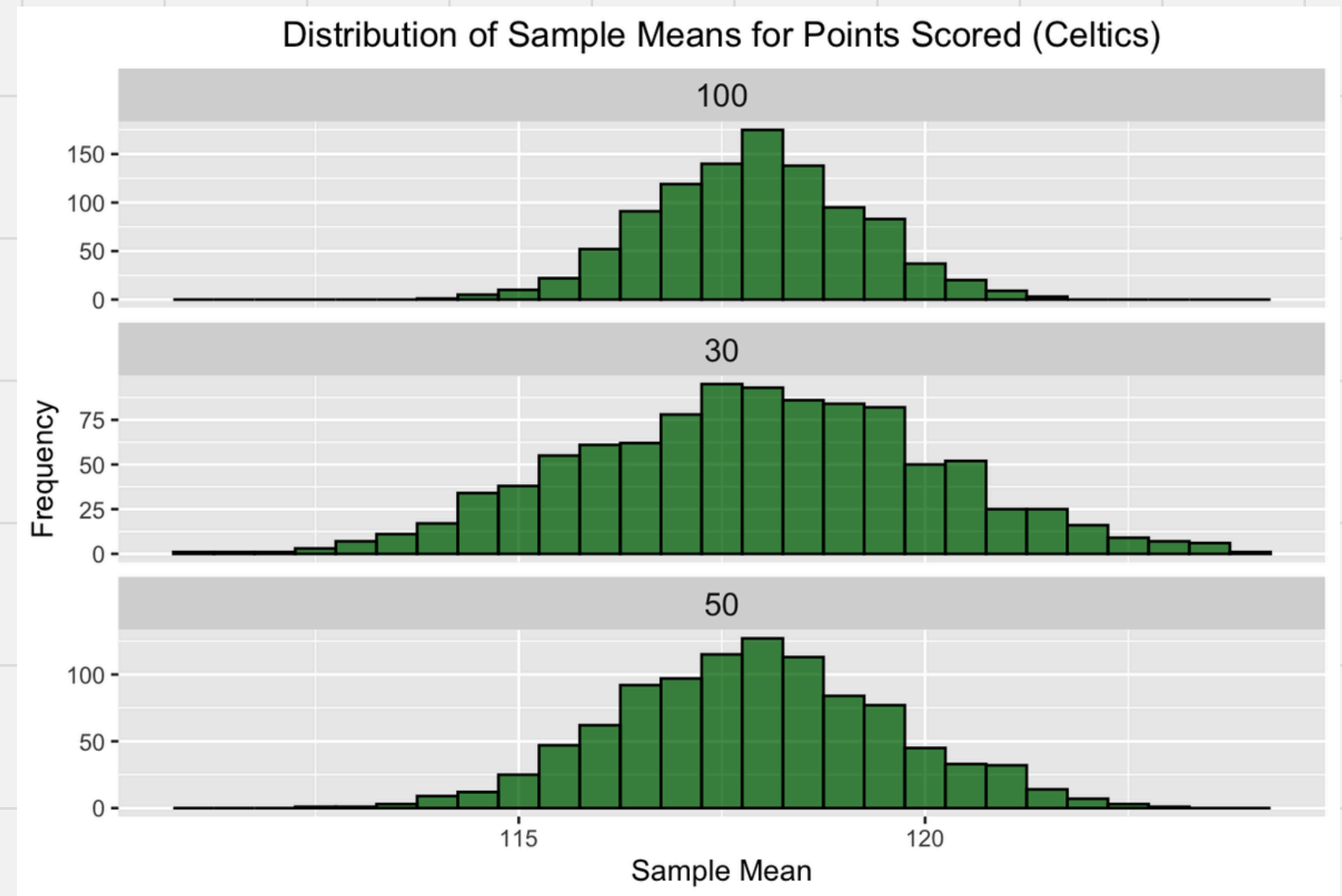
The linear regression line (in red) shows the general trend in the data, indicating the direction and strength of the relationship between the points scored and the free throws made.

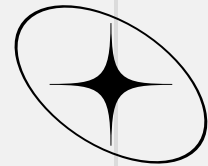




CENTRAL LIMIT THEOREM

The Central Limit Theorem states that the distribution of sample means will approximate a normal distribution as the sample size increases, regardless of the original distribution.



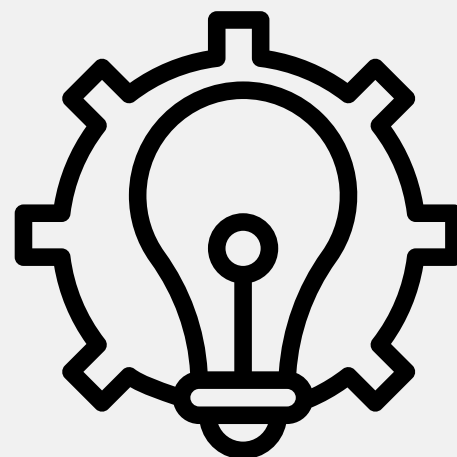


SAMPLING METHODS



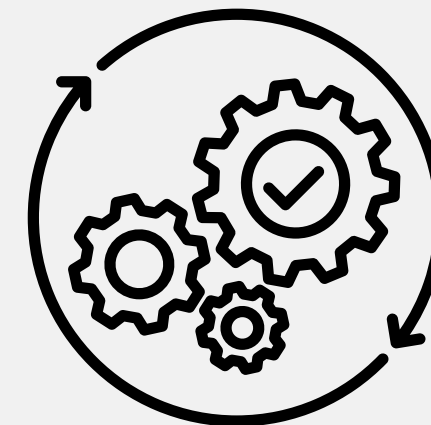
SIMPLE RANDOM

Each game in the dataset has an equal chance of being selected



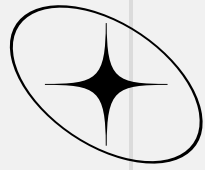
STRATIFIED

The dataset is divided into strata based on wins and losses and samples are taken from each stratum proportionally.



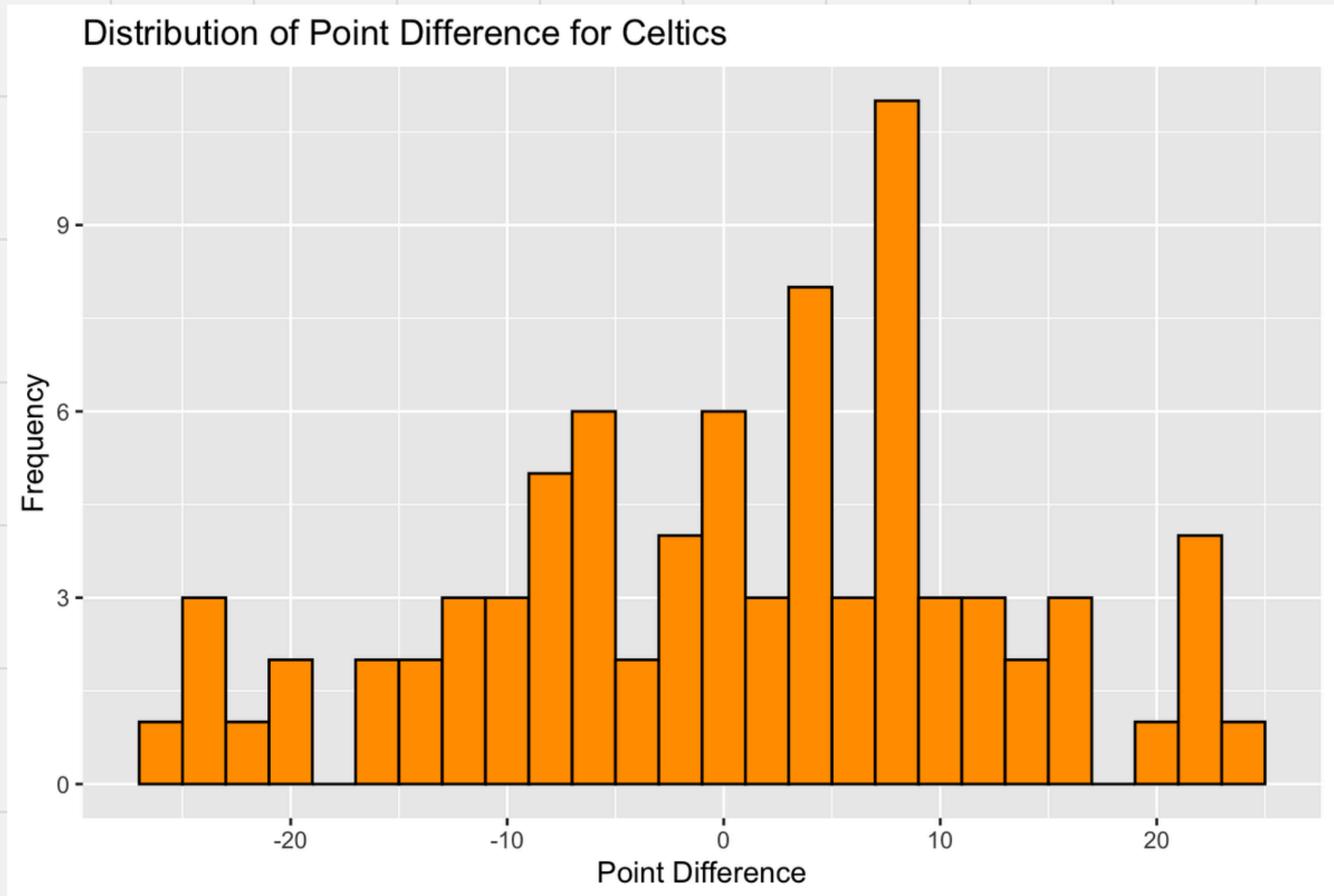
SYSTEMATIC

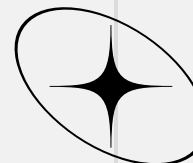
Every 10th game in the dataset is selected



ADDITIONAL ANALYSIS

The relatively symmetric distribution around 0, with a mode around 10 points, indicates that the Celtics frequently win their games, often by a margin of around 10 points.





**PERFORMANCE
ASSESSMENT**

GAME STRATEGY

**PLAYER
DEVELOPMENT**

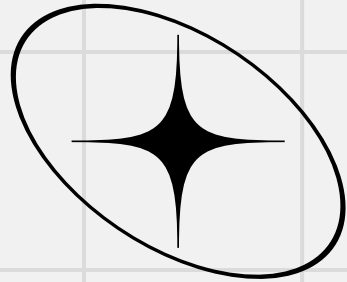
**DATA-DRIVEN
DECISIONS**

CONCLUSION



SUMMER 2024

CS 544



THANK YOU



KADER NDOYE

