

Modifying k-means clustering to optimize defensive positioning in volleyball

Naomi Consiglio
Rice University Class of 2025
B.A. in Statistics, Sport Analytics
Dr. Scott Powers Lab



RICE UNIVERSITY
Sport Analytics

Table of contents

01

Background

On volleyball and typical defensive positionings

02

Data Overview

Summary of the data available and used

03

EDA

Initial findings on attack locations

04

Adapting k-means

Modifying the algorithm to be a probabilistic model

05

Results & applications

Use results to find optimal player positioning

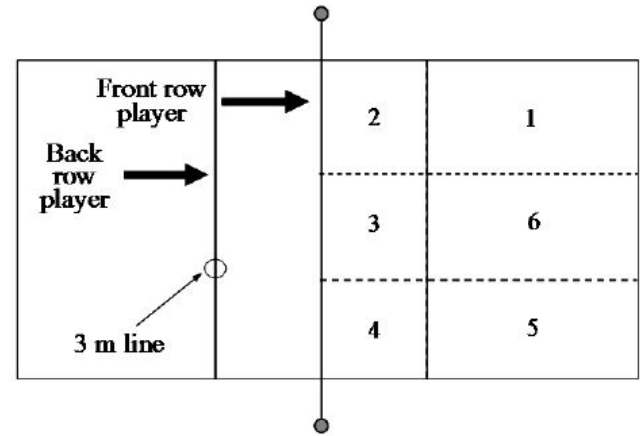
06

Next Steps

Expanding on future goals for this project

Volleyball background

- 6 players on for each team, split by zone and positions
- Positions
 - Setter
 - Outside Hitter
 - Middle Blocker
 - Rightside/Opposite Hitter
 - Defensive Specialist
 - Libero



Volleyball background, continued

- Attack
 - Hard attack
 - Offspeed/Roll
 - Tip
- Error
- Kill
- Dig
- Block



Types of defensive positioning

Base Defense

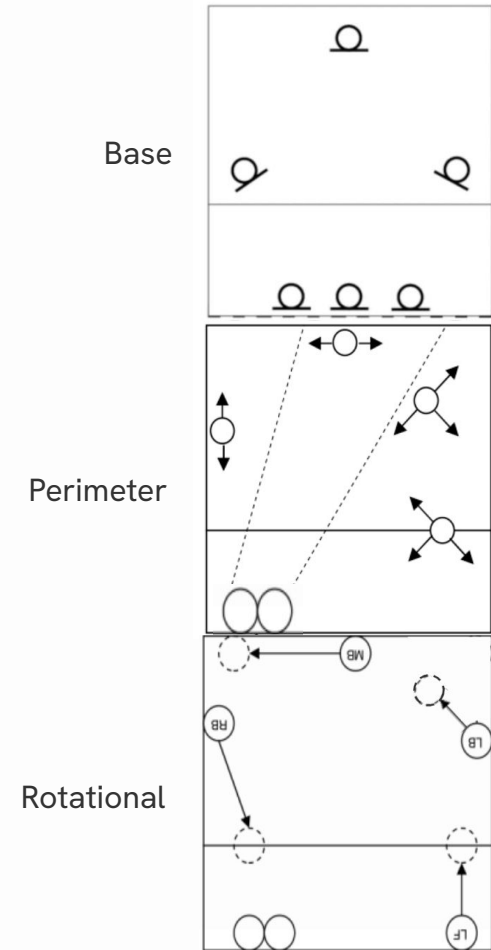
Positioning before knowing which of the opponent's attackers will attack

Perimeter Defense

Defenders surround the perimeter of the court under the idea that moving forward to the center of the court is easier than moving backwards

Rotational Defense

Assumes the blockers will cover the necessary area of the court, everyone else shifts around it



**What is the best
defensive
positioning?**

Data overview

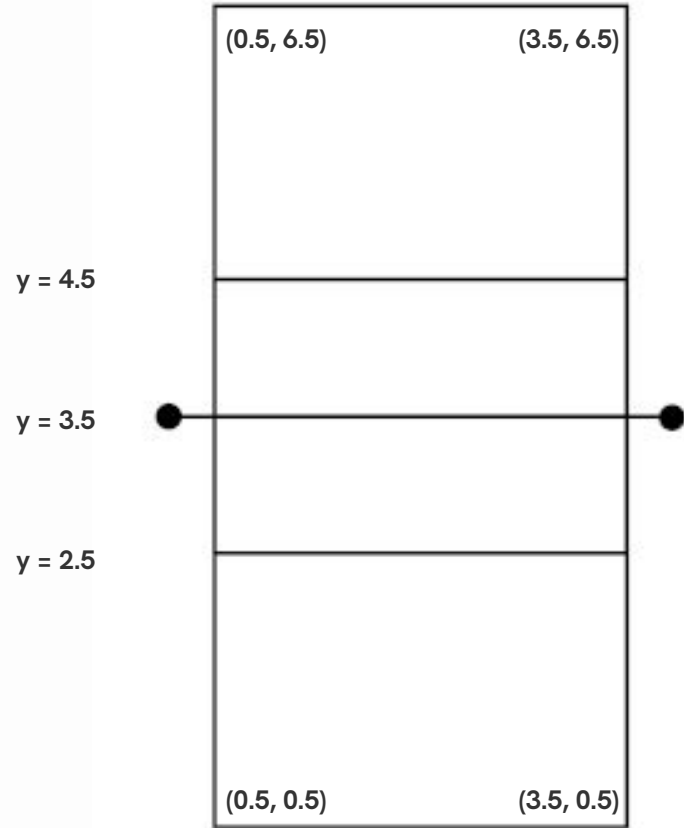
- VolleyStation manual live stat tracking
 - Player
 - Skill
 - Evaluation code
 - Skill sub-type
 - Start and end coordinates (x, y)
- Volleymetrics
 - Platform to collect video/stats of other teams
- All touches from NCAA Division I Rice University volleyball matches during the 2023 season
 - 3,771 attacks: 1,997 from outside, 1,225 from opposite, and 549 from middle
 - Attacks with significant block touches removed

The screenshot displays a live video feed of a volleyball match at Rice University, with the 'RICE OWLS' logo visible on the backboard. Below the video, a detailed stat tracking interface is shown. The interface includes a video player with a progress bar at 1:17:39 and a 1x speed setting. Below the video, there are tabs for S1, S2, and S3. The S1 tab is active, showing a list of player statistics and a scoreboard. The scoreboard indicates Rice is leading 22-13. The player list includes names and their corresponding statistics, such as K. Patel (1), G. Mansfield (2), K. Mateo (3), K. Knobbe (4), N. McCardell (5), C. Love (6), S. Hill (7), L. Ogunlana (8), T. John (9), and D. Harris (11). The interface also features a 'Shirt number' and 'Position' filter, and a 'Pts' button at the bottom.

Player	Stats
K. Patel	1
G. Mansfield	2
K. Mateo	3
K. Knobbe	4
N. McCardell	5
C. Love	6
S. Hill	7
L. Ogunlana	8
T. John	9
D. Harris	11

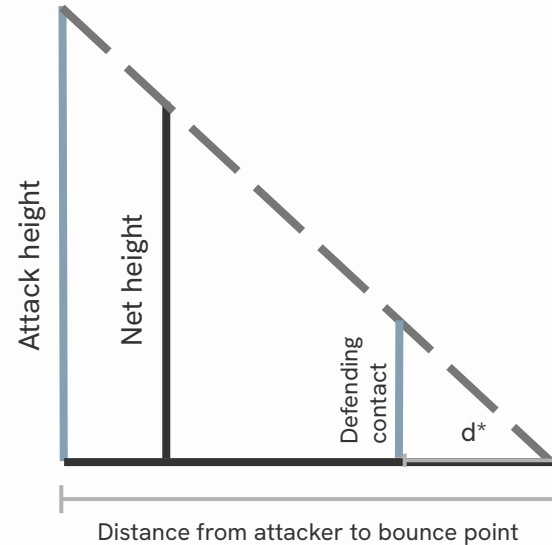
Coordinate system

- 1 coordinate = 10 feet
- Bottom baseline $y = 0.5$
- Top baseline $y = 6.5$
- Left sideline $x = 0.5$
- Right sideline $x = 3.5$
- Net $y = 3.5$
- Ten foot lines $y = 2.5, 4.5$

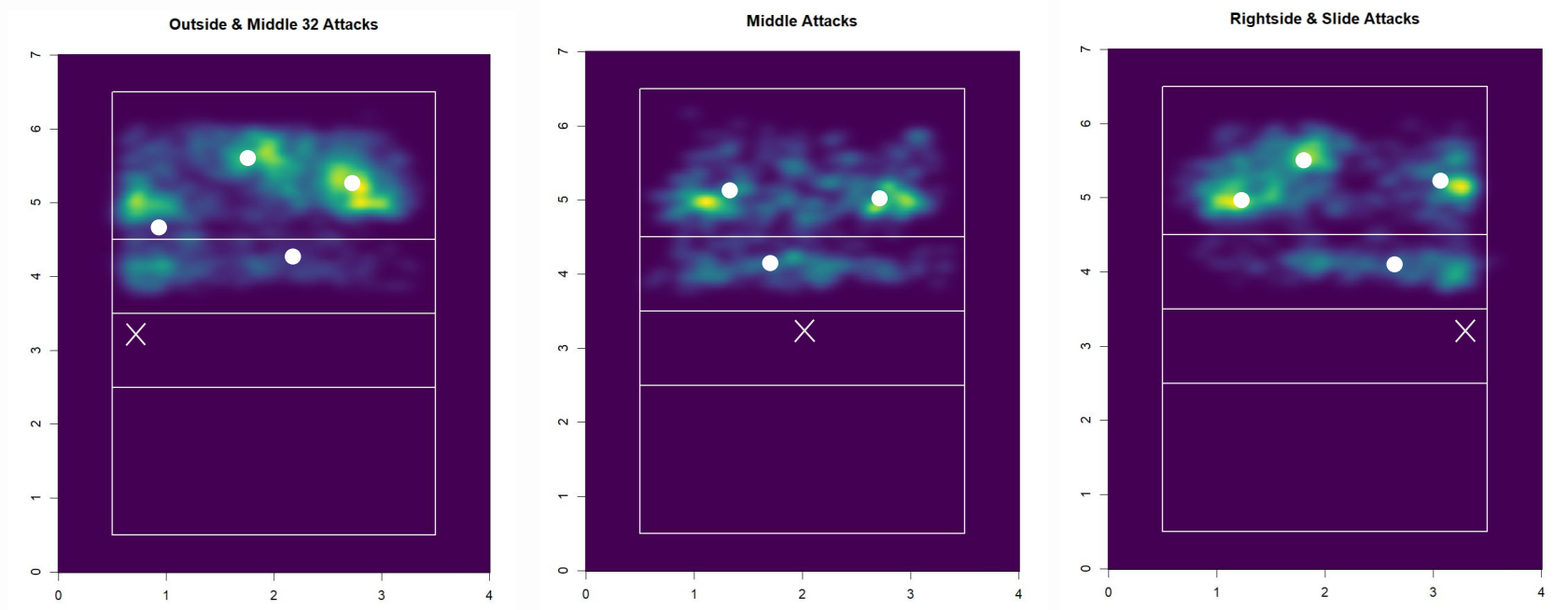


Adjustment for bounced attacks

- An attack that was not touched by a defender, including errors and kills
- Standardize attacks to have the end coordinates be where a defender should be located
- Height of attack - 100 in
 - (1 ft above the net)
- Height of defender - 30 in



Exploratory Data Analysis – k means clustering



K-means clustering vs. Adaptive model

$$\text{minimize}_{c_i} \sum_{i=1}^n ((x_i - x_{c_i})^2 + (y_i - y_{c_i})^2)$$

- Finds mean point that minimizes the distance of the nearby observations
- Found 4 means, signifying where the four spots on the court the most of each attack goes

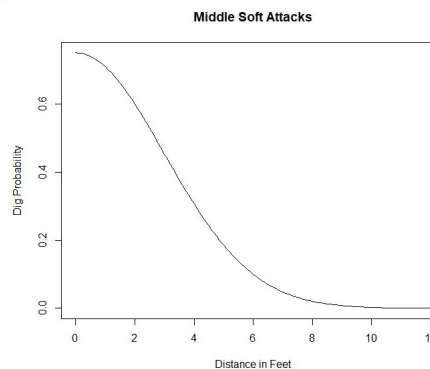
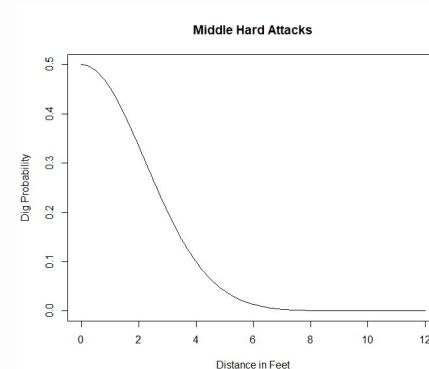
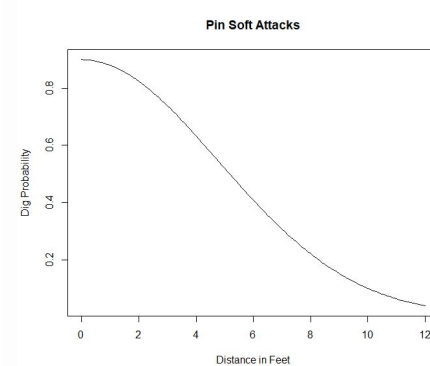
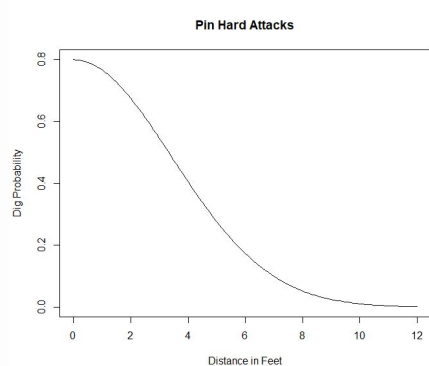
$$\text{minimize}_{c_i} \sum_{i=1}^n (P(Z_i = 1 | x_i, y_i, x_{c_i}, y_{c_i}))$$

- Minimizes the probability of the opponent getting a kill (Z_i)
- Finds the mean coordinate points that minimizes this probability
- This will tell us where the players should be on the court

Objective Function

$$\text{dig probability} = \alpha \cdot \exp(\beta \cdot \text{distance}^2)$$

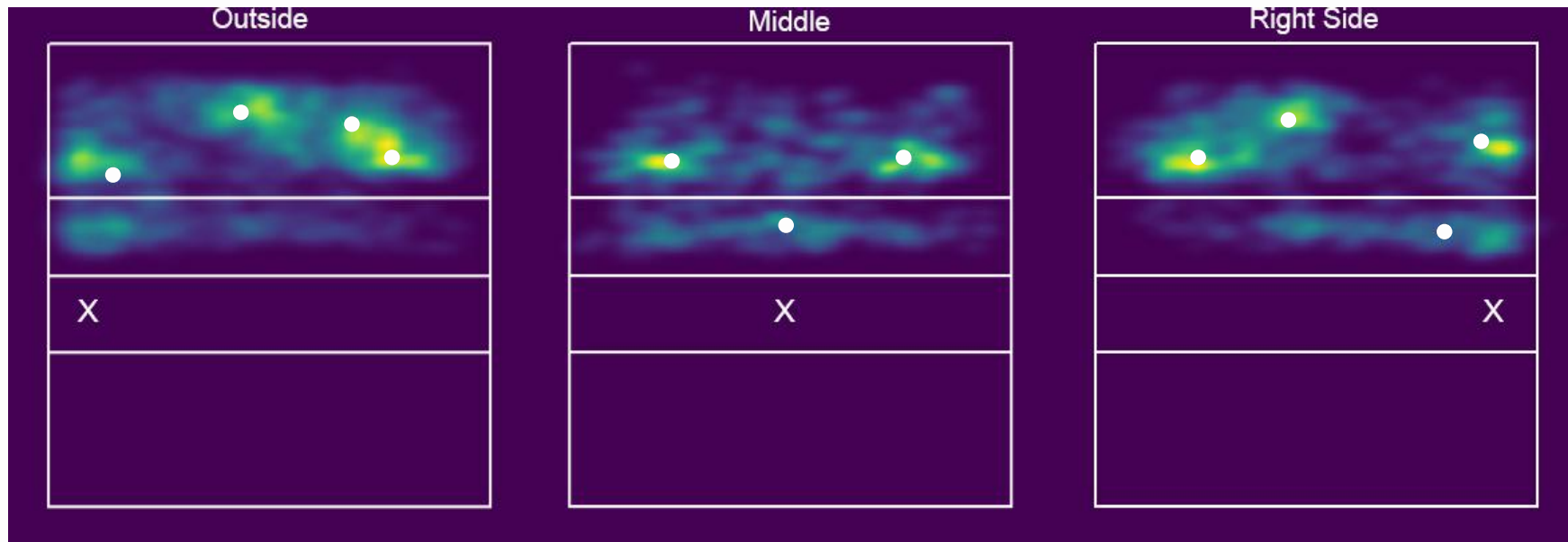
- Calculates the probability of a successful dig based on the distance between the defender and the anticipated ball landing point
- Domain knowledge and video analysis to derive α and β



Adapting k-means algorithm

1. Initialize cluster centers randomly
2. Assign each observation to the nearest center
 - a. Directly proportional to dig probability
 - b. Alternates until convergence
3. **Gradient ascent used to maximize the objective function**
 - a. Calculated the gradient of the dig probability with respect to defender positions
 - b. Iteratively update positions to increase dig probability
 - Updates until convergence

Updated results



4 defenders, 2 blockers

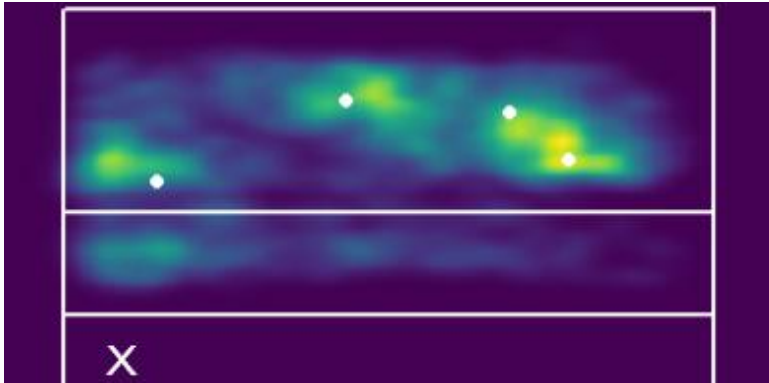
3 guaranteed defenders

4 defenders, 2 blockers

Comparison to traditional methods

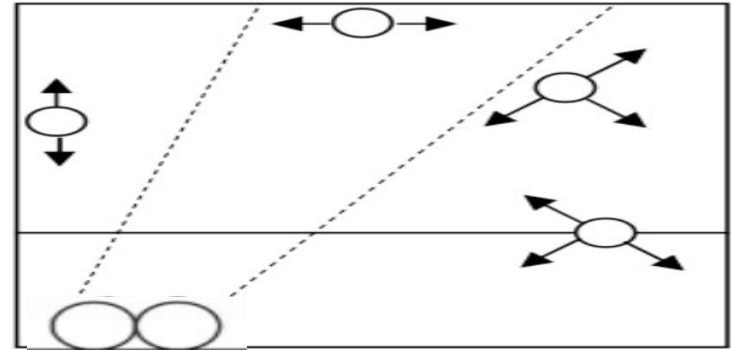
Adapted k-means result

- Pulls the offside blocker back row
- Similar to standard k-means otherwise



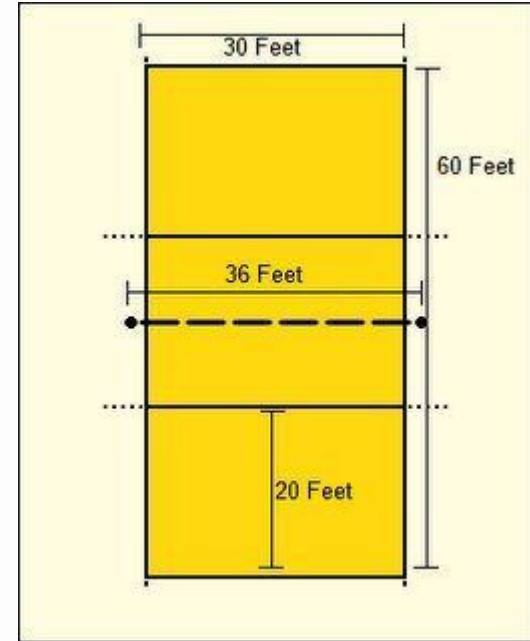
Perimeter Defense

- Closest traditional method to results
- Has players closer to side/endlines



Application

- Can easily translate coordinates to feet
 - Useful information for coaches and players
- Position players slightly deeper
- On middle attacks, middle defender move forward to cover tips over the block
- Reinforced the strategy of side defenders covering shorter areas and the middle defender managing deeper zones



Future steps

Continue testing coordinate data

- Continue testing to see if the coordinates make sense with different types of touches, including block touches

Refine player mobility information

- Currently assumes equal range of motion forward, backward, and sideways
- Adapt player range to resemble more of an oval shape

Expand data set

- Run analyses on specific teams to support scouting reports
- Apply analysis across all NCAA Division I volleyball teams for general information

Thank you!

Thank you to Dr. Scott Powers, members of the Powers lab, and Rice University Sport Analytics Department



RICE UNIVERSITY
Sport Analytics

LinkedIn:

www.linkedin.com/in/naomi-consiglio