

Geographic proximity to strong competitors and team performance in high school water polo

Ashton Jung¹, Young H. Bae²

¹ Troy High School, Fullerton, California

² Department of Marketing, Penn State University, Greater Allegheny, Pennsylvania

SUMMARY

Geographic proximity may influence competitive outcomes in youth sports because nearby teams often compete within shared regional environments and may encounter many of the same opponents. However, the spatial aspect of team performance has rarely been quantitatively examined. In this study, we investigated whether season-record performance in high school water polo exhibited spatial clustering across teams. We hypothesized that teams located near high-performing high school water polo programs would demonstrate similar performance outcomes, resulting in positive spatial autocorrelation. We collected publicly available team performance data for the 2024-2025 season from U.S. high school boys' water polo teams at the national level and within California. We analyzed these data alongside geographic location information and applied Moran's I , a spatial autocorrelation statistic that measures whether nearby teams tend to have more similar or more dissimilar performance outcomes than expected by chance. Contrary to our hypothesis, we found no evidence of positive spatial clustering in team performance. Instead, the results revealed weak but statistically significant spatial dispersion at both the national level and within California, a region with high team density and competitive intensity. Overall, the 2024–2025 season showed that proximity to strong high school water polo programs did not correspond to localized clusters of similar team performance. Our study highlights the importance of empirically testing intuitive assumptions about competitive environments and demonstrates how spatial statistical methods can be applied to evaluate performance structure in youth sports.

INTRODUCTION

A range of interrelated factors, including athlete skill, coaching quality, training opportunities, and access to institutional resources, shape high school athletic performance (1). Prior research in youth sports emphasizes that athletic development during adolescence depends not only on individual ability but also on the broader environments in which athletes train and compete (1, 2). These contextual factors play an important role in shaping performance outcomes. In addition to internal team characteristics, the competitive environment surrounding a team may influence performance

outcomes. Research on deliberate practice suggests that repeated engagement in challenging, structured activities can contribute to skill development (3). Extending this reasoning to competitive settings, teams that regularly encounter strong opponents might have greater opportunities for adaptation and learning. We therefore treated the possibility that stronger local competitive environments are associated with higher team performance as a testable assumption rather than an established effect.

Geographic location may shape competitive environments in youth sports. Teams located near strong high school water polo programs may encounter higher levels of competition through regular-season play, tournaments, and shared athletic contexts, potentially increasing competitive intensity and opportunities for skill development (1, 3). Under this assumption, we expected teams operating within the same geographic area to exhibit similar performance outcomes.

Despite the intuitive appeal of this idea, it remains unclear whether team performance in youth sports is spatially correlated rather than randomly distributed. High school boys' water polo provides a useful setting for examining the spatial organization of athletic performance. The sport has an uneven geographic distribution of teams, with a high concentration in certain regions—particularly California—and relatively sparse representation elsewhere. Publicly available MaxPreps team listings and season records provide information on teams' wins, losses, and draws (6). Rather than using the published MaxPreps ranking as the analytic outcome, we calculated each team's winning rate from the MaxPreps season-record data and used it to construct a transparent performance measure. This approach allowed us to compare teams using a reproducible season-record measure of performance.

Spatial autocorrelation refers to the degree to which observations located near one another have similar or dissimilar values. Spatial statistical methods provide a formal framework for evaluating this question by quantifying whether nearby entities exhibit more similar outcomes than expected by chance. Measures of spatial autocorrelation, such as Moran's I , allow for explicit hypothesis testing of geographic clustering in outcomes (4, 5). In the context of this study, positive spatial autocorrelation would mean that nearby teams tend to have similar performance levels, while negative spatial autocorrelation would mean that nearby teams tend to

have different performance levels.

Based on this reasoning, we hypothesized that season-record team performance in U.S. high school boys' water polo would exhibit positive spatial autocorrelation under a k -nearest neighbors spatial structure, such that geographically proximate teams would demonstrate statistically similar performance outcomes. However, our results did not support this hypothesis and instead indicated weak spatial dispersion in team performance. Overall, these findings suggested that geographic proximity to strong high school water polo programs alone was not associated with localized clustering of similar team performance in the 2024–2025 season.

RESULTS

The final analytic sample included 755 U.S. high school boys' water polo teams with complete season-record ranking and geographic location data. Because California contained the largest concentration of teams in the dataset, we also analyzed a California-only subset consisting of 571 teams. The mean reversed winning-rate rank was 378.000 (SD = 218.065) in the national sample and 382.038 (SD = 212.036) in the California subset.

Spatial autocorrelation of season-record performance at the national level

To evaluate whether team performance exhibited geographic clustering, we first examined spatial autocorrelation in season-record performance across all teams in the dataset. We quantified competitive outcomes using reversed season-record rankings, such that higher values corresponded to stronger performance. We defined spatial relationships among teams using a k -nearest neighbors spatial weights matrix ($k = 8$) based on geographic distance, which balanced the representation of immediate geographic context with sufficient connectivity across regions with uneven team density.

Global Moran's I analysis revealed a small but statistically significant negative spatial autocorrelation at the national level (Moran's $I = -0.050$, permutation p -value = 0.003). We illustrated the global Moran's I results using a Moran's I scatterplot (Figure 1). We plotted each team's reversed season-record ranking against the spatial lag of neighboring teams' rankings. The fitted regression line corresponded to the Moran's I statistic and exhibited a shallow negative slope, consistent with the observed negative Moran's I value. The dispersion of points across all four quadrants and the absence of concentrated high–high or low–low groupings indicated weak spatial dispersion rather than positive geographic clustering in team performance. This suggested that teams located near one another tended to exhibit slightly dissimilar performance outcomes.

To further explore localized spatial patterns, we computed local Moran's I statistics for individual teams. A limited number of teams were associated with statistically significant local spatial relationships ($p < 0.05$); however, these associations

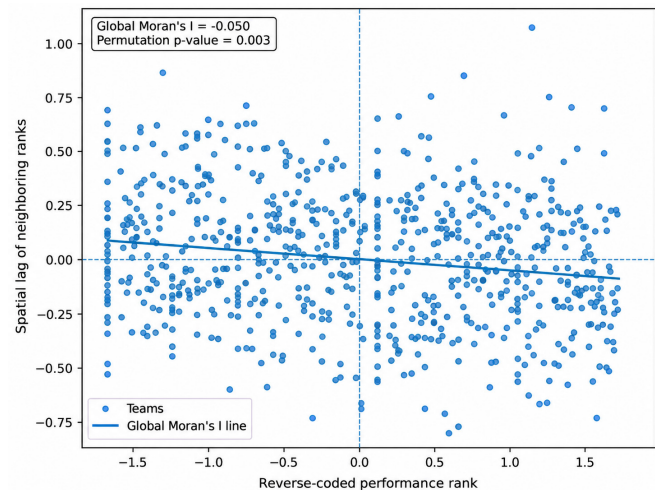


Figure 1: Global Moran's I scatterplot for national high school boys' water polo performance. Each point represents a team's reverse-coded performance rank plotted against the spatial lag of neighboring teams' ranks using a k -nearest neighbors spatial weights matrix ($k = 8$). The fitted line corresponds to global Moran's I statistic (Moran's $I = -0.050$, permutation p -value = 0.003), indicating weak negative spatial autocorrelation.

were sparse and geographically scattered. We observed both high–high and low–low local associations, along with mixed high–low and low–high patterns (Table 1). These statistically significant local associations represented a small minority of teams and included all four cluster types, indicating localized spatial heterogeneity rather than a broad pattern of regional performance clustering.

Spatial autocorrelation of performance within California

Because high school water polo teams are unevenly distributed across the U.S. with a particularly high concentration in California, we conducted a secondary analysis restricted to teams located within California. This analysis allowed us to examine whether spatial clustering emerged in a region characterized by high team density and frequent interscholastic competition.

As a descriptive overview of performance patterns within California, we mapped winning-rate-based performance scores for teams located throughout the state (Figure 2). High- and low-performing teams appeared interspersed rather than visibly concentrated into distinct geographic regions. Using the same $k = 8$ nearest-neighbors specification as in the national analysis, global Moran's I within California indicated weak but statistically significant negative spatial autocorrelation (Moran's $I = -0.040$, permutation p -value = 0.014). We illustrated this result using a Moran's I scatterplot for the California subset (Figure 3). The fitted regression line exhibited a shallow negative slope, consistent with the observed negative Moran's I value. The distribution of points indicated weak spatial dispersion rather than positive spatial clustering of season-record performance.

Cluster type	Description	Local Moran's <i>I</i> relationship	Number of teams
HH	High-performing teams near high-performing teams	Positive	10
HL	High-performing teams surrounded by low-performing neighboring teams	Negative	14
LH	Low-performing teams surrounded by higher-performing neighboring teams	Negative	16
LL	Low-performing teams near low-performing teams	Positive	4

Table 1: Statistically significant local Moran's *I* cluster types in the national analysis. The national analysis included 755 teams. Cluster types describe the relationship between each team's performance and the performance of its *k*-nearest neighboring teams. Local Moran's *I* were calculated using reverse-coded winning-rate ranks, row-standardized *k*-nearest-neighbors weights (*k* = 8), and 999 permutations. Statistical significance was assessed at $p < 0.05$.

Local Moran's *I* analysis within California identified 30 statistically significant local associations ($p < 0.05$); however, these associations represented a small minority of California teams and supported localized spatial heterogeneity rather than a broad regional pattern of similar performance. We observed isolated instances of high-performing teams located near lower-performing neighbors and vice versa.

Sensitivity analysis of neighborhood size

To evaluate the effect of neighborhood size, we conducted sensitivity analyses using *k*-nearest-neighbor values of 4, 6, 8, 10, and 12. At the national level, Moran's *I* remained negative and statistically significant across all tested specifications, ranging from -0.041 to -0.050 (permutation *p*-values = 0.002 – 0.035). Within California, Moran's *I* also remained negative across all tested *k* values, ranging from -0.023 to -0.040 ; however, statistical significance varied by neighborhood definition, with significant results observed for *k* = 8, 10, and 12 but not for *k* = 4 or 6. These findings indicated that the direction of the observed spatial pattern was consistent across alternative neighborhood sizes, although the strength of statistical evidence was more sensitive to *k* within California.

DISCUSSION

The results of this single-season analysis did not support our original hypothesis that geographically proximate high school water polo teams would exhibit similar performance outcomes. Instead, team performance showed weak negative spatial autocorrelation at both the national and California levels. Taken together, these findings suggest that geographic proximity to strong programs alone does not organize team performance into localized clusters, even in a region with high team density and competitive intensity.

One possible explanation for the absence of positive spatial clustering is that exposure to competitive opponents

does not translate uniformly into performance gains. Teams located near successful programs may compete in similarly demanding environments, but their ability to benefit from such exposure may depend on internal program characteristics, including coaching quality, training structure, athlete development pathways, and institutional support. Thus, the absence of positive spatial clustering does not imply that competitive environments are irrelevant to athletic development. Rather, geographic proximity to strong programs may capture only one aspect of the broader environment in which teams develop and compete.

The California results further support this interpretation. California has a high concentration of water polo programs and frequent interscholastic competition, making it a setting in which positive clustering might be expected if geographic proximity to strong programs were sufficient to produce similar performance outcomes. Instead, the California-restricted analysis mirrored the national results, with weak negative rather than positive spatial autocorrelation. The magnitude of this negative association was small, however, and does not provide evidence of a strong tendency for neighboring teams to perform differently. More appropriately, the results indicate that similar performance was not strongly organized geographically, even in a region with high team density and competitive intensity.

Several limitations should be considered when interpreting our results. First, we measured team performance using winning-rate-based results from a single season. Although this measure summarizes competitive outcomes, it does not capture developmental changes over time and cannot determine whether the observed spatial patterns are stable across seasons. Changes in team composition, schedules, league structures, and regional competition may alter these relationships in other years. Second, winning rate may reflect not only team quality but also schedule composition and the strength of opponents faced. Because league, conference, or section assignments are often geographically organized, nearby teams may play many of the same opponents, which could influence observed spatial patterns in season-record performance. Third, although we used a *k*-nearest neighbors approach to define spatial relationships, alternative definitions of competitive neighborhoods could yield different spatial structures. The sensitivity analyses showed that the negative direction of Moran's *I* was generally consistent across alternative values of *k*, but statistical significance within California varied by neighborhood definition. Fourth, the analysis did not directly measure potentially important program- and school-level characteristics, including coaching and training structures, school resources, community socioeconomic context, access to aquatic facilities and club water polo programs, travel resources, public or private school status, and program stability. Because we did not include these factors in the analysis, we could not directly evaluate their relationships with team performance and local competitive environments.

Future research could address these limitations by examining multiple seasons to determine whether the observed spatial patterns are stable over time. Longitudinal analyses could help distinguish persistent geographic patterns from season-specific variation in team composition, schedules, and competitive structure. Future studies could also incorporate direct measures of program, school, and community characteristics alongside geographic proximity. In addition, network-based measures based on actual opponents and competition schedules could provide a more direct representation of competitive exposure than geographic proximity alone. Using alternative measures of team performance could further clarify whether spatial patterns differ depending on how athletic success and development are defined.

These findings also have practical and methodological implications. For coaches, athletes, and school administrators, the results do not support the assumption that simply being located near strong water polo programs produces similar or improved team performance. Geographic context may matter, but the present findings indicate that proximity to successful teams alone does not define a team's competitive outcome. Spatial statistical methods therefore provide a useful way to test assumptions about competitive environments while also showing when geographically detectable patterns are limited in practical magnitude.

From a methodological perspective, these findings also highlight the importance of distinguishing between statistical significance and effect magnitude. Although Moran's I was statistically significant under the primary specification at both the national and California levels, its magnitude was small. Thus, the results indicate a statistically detectable but weak pattern of spatial dependence.

In conclusion, the 2024–2025 data did not show positive geographic clustering in winning-rate-based team performance among high school boys' water polo teams. Instead, the observed pattern showed weak negative spatial autocorrelation, including within California under the primary $k = 8$ specification. We interpret these findings as evidence from a single season rather than as a general conclusion about spatial patterns in high school water polo. More broadly, this study demonstrates the value of spatial statistics for testing assumptions about competitive environments and shows that geographic proximity to strong programs does not necessarily translate into similar patterns of team performance.

MATERIALS AND METHODS

Data sources and study sample

Data were obtained exclusively from publicly available sources. MaxPreps was used to identify U.S. high school boys' water polo teams and obtain publicly available season-record information, including wins, losses, and draws (6). The published MaxPreps ranking was retained as a source variable for identifying teams but was not used as the primary performance outcome. Institutional information, including

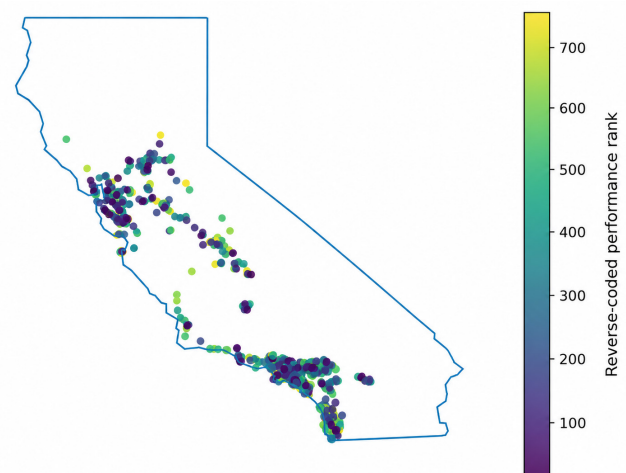


Figure 2: Spatial distribution of high school boys' water polo performance across California. Each point represents the geographic location of a high school boys' water polo team in California. Point color indicates reverse-coded performance rank, where higher values correspond to stronger competitive performance. High- and low-performing teams appear interspersed throughout the state with no clear regional clusters.

school name, location, and public or private classification, was cross-referenced using publicly available Collegiate Water Polo Association materials and official school athletic department websites when necessary.

The initial MaxPreps season-record ranking list included 756 high school boys' water polo teams for the 2024–2025 season. Teams were included in the analysis only if complete season-record and geographic location data could be verified. One team was excluded during data cleaning because complete ranking and geographic information could not be verified. After applying these exclusion criteria, the final analytic dataset consisted of 755 high school boys' water polo teams, with the dataset finalized as of January 23, 2025.

Geographic data processing and variable construction

Each school's geographic location was determined using its publicly listed street address. Addresses were converted to latitude and longitude coordinates through geocoding and manually verified using Google Maps (7). These coordinates were used to compute geographic distances between teams. Team performance was operationalized using a researcher-defined, winning-rate-based performance score. Total games played were calculated as the sum of wins, losses, and draws. Winning rate was calculated as wins divided by total games played. Teams were then ranked by winning rate, with higher winning rates assigned stronger ranks. Tied teams were assigned average ranks. These ranks were reverse-coded so that higher numerical values represented stronger season-record performance. The resulting reverse-coded winning-rate rank was used as the dependent variable in all global and local Moran's I analyses.

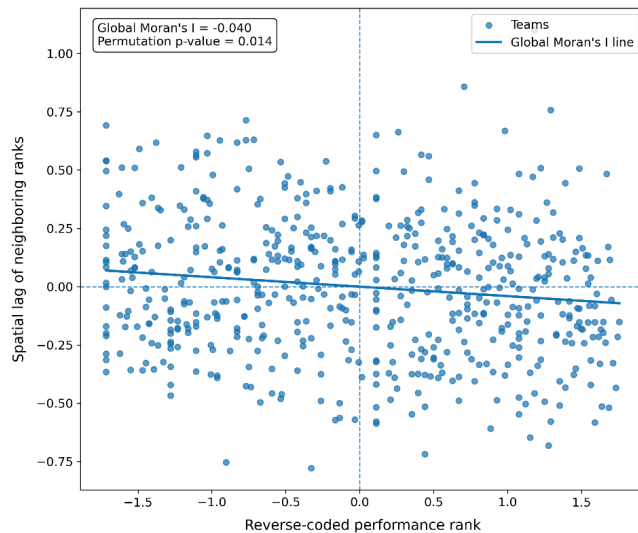


Figure 3: Global Moran's I scatterplot for California high school boys' water polo performance. Each point represents a California team's reverse-coded performance rank plotted against the spatial lag of neighboring teams' ranks using the same k -nearest neighbors spatial weights matrix ($k = 8$). The fitted line corresponds to the global Moran's I statistic (Moran's $I = -0.040$, permutation p -value = 0.014), indicating weak negative spatial autocorrelation.

Spatial weights and autocorrelation analysis

Spatial relationships among teams were defined using a k -nearest neighbors approach based on geographic distance. For each team, the k geographically closest teams were identified using Euclidean distances calculated from latitude and longitude coordinates. A value of $k = 8$ was selected to capture local competitive neighborhoods while maintaining sufficient connectivity across the dataset. To assess whether the findings were sensitive to this choice, the global Moran's I analyses were repeated using alternative k -nearest-neighbor specifications of $k = 4, 6, 10$, and 12 . The resulting spatial weights matrix was row-standardized prior to analysis. Because teams were unevenly distributed geographically, the spatial weights matrix contained a small number of disconnected components but this structure is permissible for Moran's I analysis and does not invalidate permutation-based inference.

Global spatial autocorrelation in team performance was assessed using Moran's I statistic (8). Spatial autocorrelation describes whether observations located near one another tend to have similar or dissimilar values. In this study, Moran's I was used to test whether geographically nearby teams had similar reversed ranking values, which would indicate spatial clustering, or dissimilar reversed ranking values, which would indicate spatial dispersion. Moran's I was calculated as:

$$I = \left(\frac{N}{W} \right) \times \left[\frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \right] \quad (\text{Equation 1})$$

where N is the number of teams, and x_i and x_j are the reversed ranking values for teams i and j , $\bar{x}$ is the mean reversed ranking

value, is the spatial weight between teams i and j , and W is the sum of all spatial weights.

Positive Moran's I values indicate that nearby teams tend to have similar performance outcomes, negative values indicate that nearby teams tend to have dissimilar performance outcomes, and values near zero indicate little or no spatial autocorrelation (4). Moran's I was computed for reversed rankings at the national level and for a subset of teams located within California to examine spatial patterns in a region with high team density. Statistical significance was evaluated using permutation tests with 999 random permutations. Local Moran's I statistics were also computed to identify localized spatial associations between individual teams and their geographic neighbors.

Software and reproducibility

All data processing and analyses were conducted in Python in the Google Colab environment. Spatial analyses were performed using open-source libraries from the PySAL ecosystem, including libpysal for spatial weights construction and esda for Moran's I computation, along with GeoPandas for spatial data handling and visualization. Custom scripts were used for data cleaning, variable construction, spatial analysis, and figure generation. The analysis notebook, final analytic dataset, output tables, figures, and required software packages are available in [a public GitHub repository](#).

Received: January 19, 2026

Accepted: June 30, 2026

Published: October 2, 2026

REFERENCES

- Baker, Joseph, and Sean Horton. "A Review of Primary and Secondary Influences on Sport Expertise." *High Ability Studies*, vol. 15, no. 2, 1 Dec. 2004, pp. 211–228. <https://doi.org/10.1080/1359813042000314781>
- Hrysomallis, Con. "Balance Ability and Athletic Performance." *Sports Medicine*, vol. 41, no. 3, Mar. 2011, pp. 221–232. <https://doi.org/10.2165/11538560-000000000-00000>
- Ericsson, K. Anders, et al. "The Role of Deliberate Practice in the Acquisition of Expert Performance." *Psychological Review*, vol. 100, no. 3, Jul. 1993, pp. 363. <http://dx.doi.org/10.1037/0033-295X.100.3.363>
- Anselin, Luc. "Spatial Dependence in Regression Error Terms." *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, 1988, pp. 100–118. https://doi.org/10.1007/978-94-015-7799-1_8
- Rey, Sergio J., and Brett D. Montouri. "US Regional Income Convergence: A Spatial Econometric Perspective." *Regional Studies*, vol. 33, no. 2, 1 Apr. 1999, pp. 143–156. <https://doi.org/10.1080/00343409950122945>
- MaxPreps. High School Boys' Water Polo Rankings. <https://www.maxpreps.com>. Accessed Jan. 23, 2025.
- Google Maps. <https://maps.google.com>. Accessed 23

Jan. 2025.

8. Wakefield, Jon. "Disease Mapping and Spatial Regression with Count Data." *Biostatistics*, vol. 8, no. 2, 1 Apr. 2007, pp. 158–183.

Copyright: © 2026 Jung and Bae. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.