Cartography M.Sc.

Master thesis

# User evaluation of the effectiveness of geoprocessing tools for football data visualization

Joel Salazar

Technical University of Munich

TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology

TECHNISCHE UNIVERSITÄT DRESDEN

ITC UNIVERSITY OF TWENTE.

2023

# User evaluation of the effectiveness of geoprocessing tools for football data visualization

submitted for the academic degree of Master of Science (M.Sc.)
conducted at the Department of Aerospace and Geodesy
Technical University of Munich


Author: Joel Salazar
Study course: Cartography M.Sc.
Supervisor: Nianhua Liu Mag. M.Sc. (TUM)
Reviewer: Gustavo Adolfo García Chapeton PhD (University of Twente)

Chair of the Thesis
Assessment Board: Prof. Dr. Liqiu Meng



Date of submission: 07.09.2023

## Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

**"User evaluation of the effectiveness of geoprocessing tools for football data visualization"**

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Munich, 07.09.2023 Joel Salazar

# Acknowledgements

Thanks to everyone who helped filling out the survey; I hope you had some fun.

A special thanks to Juliane Cron for being such an amazing soul who helped me along the Cartographic journey.

I love you, mom. Thank you, Neri and Nicole, for always being by my side.

Thanks to all my new friends I made along this journey, I enjoyed every time we spent together.

# Abstract

Geovisualizations offer a powerful way to capture the dynamic nature of football games through spatiotemporal analysis and visualization (Andrienko et al., 2021; Kotzbek & Kainz, 2014), enhancing the perception and understanding of tactics in football. Geoprocessing tools such as Nearest Distance calculation, Voronoi diagrams, and Convex Hull have been used for data analysis rather than visualization. This study aims to address this gap by answering the question: How *effective are geoprocessing tools on football understanding in users with football knowledge and without football knowledge?* Four animations were designed for the evaluation (one for each geoprocessing tool and a raw animation). The animations show a counter-attack football event of 30 seconds from a real game. The data used is from Metrica Sports (Metrica Sports, n.d.). Based on the literature review, I selected six parameters to evaluate users' understanding of football tactics: *playing formation, dominant region, playing space*, *attacker-defender distance*, the *relative distance for a defender to intercept a shot*, and the *distance between teammates*. The tool for the evaluation was an online survey. The survey collected data from 109 participants. Participants with football knowledge represent 64% of the sample (n=70). The sample remaining 36% (n=39) represent participants without football knowledge. The effectiveness criteria were the difference between correct answers from the animation with the geoprocessing tool and the raw animation. Results show that the Nearest Distance calculation is more effective in visualizing the *playing formation* of a team and the *attacker-defender distance* of players. The Convex Hull is more effective for visualizing the *playing space* of a team. The Nearest Distance calculation is ineffective for visualizing a defender's relative distance *to intercept a shot*. The remaining parameters, such as *dominant region, and the distance between teammates*, show variations in results, either positive or negative effectiveness, in each of the groups. These results contribute to close the gap between cartography and football data analysis and serve as a reference for further research in cartographic football visualization.

# Contents

# 1. Introduction

## 1.1. Motivation and problem statement

Football, also known as soccer, is a sport played by two teams with eleven players each, within a 90-minute time limit. During the game, both teams engage in competitive and cooperative movements, creating dynamic spatial and temporal patterns (Stein et al., 2018). A football match can be seen as a complex spatiotemporal framework in which various phenomena and objects are interrelated, including the ball, the players, referee, and referee assistants. The position of the players and the ball are unknown in advance, but recent progress in football data collection has led to the gathering of vast amounts of data after a professional football match (Kotzbek & Kainz, 2014).

The availability of this data has created new opportunities for providing data-driven insights into the game, attracting the attention of various stakeholders such as coaches, players, football analysts, journalists, researchers, and the general public (Andrienko et al., 2021). Analysing football data can help identify patterns on both individual and team levels, providing insights into improving performance as well as understanding of the game itself (Perin et al., 2013; Wu et al., 2019).

Geovisualizations offer a powerful way to capture the dynamic nature of football games through spatiotemporal analysis and visualization (Andrienko et al., 2021; Kotzbek & Kainz, 2014), enhancing the perception and understanding of tactics in football. Current research on football data visualization has primarily focused on developing interfaces for handling and generating data with a data analytics orientation (Perin et al., 2013; Stein et al., 2018; Wu et al., 2019). Geoprocessing tools such as Nearest Distance calculation, Voronoi diagrams and Convex Hull have been used for data analysis rather than for visualization purposes. This study aims to address this gap by evaluating how users that know and do not know about football understand specific parameters of football tactics. This study seeks to provide valuable insights for the future development of football geovisualization.

## 1.2. Research identification

The main objective of this thesis is:

- To evaluate the effectiveness of three geoprocessing tools (Nearest Distance calculation, Voronoi diagram, and Convex Hull) for visualizing football tracking data.

The main objective consists of three sub-objectives:

- To generate three geovisualizations based on open football tracking data using the geoprocessing tools considered.
- To design a user experiment to evaluate the effectiveness of the geovisualizations within users with and without football knowledge.

- To find football data analysis parameters to link previous research with the effectiveness evaluation.

The following research question is used to address the main objective:

> **RQ1.** How *effective are geoprocessing tools on football understanding in users with football knowledge and without football knowledge?*

This main question will be addressed by the next sub question to frame the evaluation design:

> ***RQ2.*** *Which parameters of football data analysis can be considered to evaluate the three geoprocessing tools?*

This research helps a step forward into understanding user's perception of football geovisualizations, as well as the usage of football data into the cartographic field. An online survey is used for the evaluation phase. The scope of this research will be determined by the number of users who participate in the evaluation survey. The research is not focused on gathering or mining football data, nor is it aimed at proposing new types of geovisualizations for football data.

## 1.3. Structure of the thesis

This research is structured in six chapters. The Theoretical framework aims to give a background of literature around football data handling and visualization, as well as describing geoprocessing tools used to analyse football data and create a link between both concepts, football data and geoprocessing tools for the overall research. The third chapter, Methodology, describes the process of this work that was divided into two parts. First, the description of the football dataset used, and the generation of football animations based on the three geoprocessing tools that are analysed: Nearest Distance calculation, Voronoi diagrams, and Convex Hull. Second, the description of the design of the survey as a tool to evaluate the understanding of the animations from users.

Chapter 4 presents the results according to the structure of the description of the geoprocessing tools in the previous chapters. It starts with the results regarding the question about point pattern distribution or the formation of teams on the pitch. It follows the results for the three questions that aims to test the Nearest Distance animation. The Voronoi diagram is next with its own data description and finally, the Convex Hull animation, with the results for the question that was used to test it.

The Discussion of the results is presented in Chapter 5, which analyses the results and answers the research question. This research ends with Chapter 6 as the Conclusion with an outlook of the work that has been done and an exposure of limitations and ideas for further research.

# 2. Theoretical framework

## 2.1. Introduction

The goal of this chapter is to give an overview of literature related to football data and geovisualizations, aiming to create a link between both concepts. I will start by describing football data and its characteristics as a spatiotemporal dataset. Then, three geoprocessing tools are described: nearest distance calculation, Voronoi diagrams, and convex hull. These descriptions will be used to frame the theoretical background to further analyze the mentioned geoprocessing tools with the selected football geodataset.

## 2.2. Football data

Advanced tracking technologies in sports, especially in professional football, allows us to gather detailed data on player movements during matches. After each game, this large amount of data is used for post-match analysis, helping coaches, players, and analysts gain a better understanding of the game. The work of Perin et al. (2018) makes a categorization of sports data that is available after a game. The research identifies three categories of sports data: box score data (data containing statistical summaries of a sport event such as a game), tracking data (data about in-game actions and trajectories), and meta-data (data about the sport and its participants but not necessarily a given game).

According to Pappalardo et al. (2019), football data is collected through three sources. Events during a match are described as soccer-logs. Trajectories of players and the ball during a match, gathered through video-tracking. Geographic position of players and the ball, collected through GPS devices. The data gathered contains detailed location of the players on the pitch. The location of the players makes it a robust spatial dataset that can be analyzed with geoprocessing tools.

Football spatial data is available in two different types: tracking data and event data. Tracking data continuously records the position of the players and the ball along a match. Temporal resolution is highly detailed and can reach up to 25 records per second and a spatial resolution of around 30cm (Kotzbek & Kainz, 2014). Event data adds an attribute linked to a certain event during the match, which can be a pass, goal, shot, etc.

## 2.3. Visualizing and handling football data

The interest in tracking data in sports has been evolving within Cartography and Geographic Information Science. Demaj (2013) describes an approach to visualize tennis spatiotemporal data with a Geographic Information System (GIS), using geospatial data analysis. The author suggests the potential in GIS tools for a better understanding of patterns of movements using spatiotemporal analysis.

With the vast amount of spatiotemporal data available after a football match, the analysis of event data and trajectories has derived in the development of visual interfaces that can provide insights or identify patterns, so to understand the game and provide meaningful performance information. Perin et al. (2013) presented a visualization interface to support soccer analysts in exploring data and communicating insights along a game. The study included an approach to usability evaluation, working along with soccer analysts.

Different methods have been proposed in the way of handling football data. For instance, Bialkowski et al. (2014) presented a method which can conduct both individual player and team analysis, stating the issue of aligning a player position over time, due to the dynamic, continuous, and multiplayer nature of football. Andrienko et al. (2017) proposed a computational approach to detecting and quantifying the relationships of pressure emerging during a game.

Visualization design alternatives are also proposed. Stein et al. (2018) proposes a visual analytics system that uses video recordings with an abstract visualization of underlying trajectory data. Wu et al. (2019) proposes a system to represent changes in team formation, allowing analysts to visually analyze the evolution of formations changing along the game and tracking the spatial flow of players within formations over time.

Post-match analysis in football is mainly done using video, as it is the medium coaches and players use and prefer most. Bradley et al. (2020) provides insights and an approach to combine data visualization in combination with video, that can translate to an animation visualization of specific moments along a game. To analyze trajectories of multiple simultaneously moving objects, such as football players during a game, Andrienko et al. (2021) proposes an approach to extract and understand the general patterns of coordinated movement in different classes of situations along a game.

In a more specific focus to football data, Liu et al. (2022) analyzed outcomes of visualizations with spatiotemporal framework such as pass map, heat map and positioning map, which intend to have a better visualization and tactical instructions effect. The authors highlight the possibilities of football animated visualizations and describe advantages and development trends. Moreover, throughout the thesis work of Liu (2022), a set of football visualizations and animations are analyzed, compiled, and compared. Steps for visualizing football data through different technologies are described and insights are shared.

## 2.4. Geovisualizations

- *Football pitch as a map*

According to Kotzbek and Kainz (2014), football can be seen as a spatiotemporal framework where various objects such as the football pitch, goals, players, the ball, and the referee and assistants are interconnected in space and time. The dynamic components of this framework consist of the ball, players, and referees. These entities are constantly in motion, interacting with one another in a spatiotemporal context, creating a dynamic and ever-changing gameplay. The events occurring during a match, such as passing, tackling, shots, and more, are part of the phenomenon involving the players. However, the analysis post-match usually does not consider data related to pitch conditions or environmental conditions, such as weather.

Spatially, the interaction between and among objects happens within a spatial frame (the pitch) that in association football has a size of 105 by 68 meters (according to the International Football Association Board – IFAB). Within the boundaries of the pitch, there is a division into certain areas, for instance, half the size of the entire pitch for each team. These divisions are relevant for the rules of the game. Temporally, football interactions occur within a total time frame of 90 minutes, split into two halves of 45 minutes each, with the possibility of additional time depending on the game's course. The analysis of football data considers the spatial distribution of players on the pitch, which is determined by the interaction behavior happening in two levels: individual player behavior and team behavior (Fonseca et al., 2013).

- *Geoprocessing tools*

Post-match football analysts try to understand spatial point patterns during a game. Techniques such as convex hull, vertical stretch and centroid position are used to describe team tactical or organizational behavior through visualizing spatial point patterns (Fonseca et al., 2013). Coito et al. (2022) grouped five categories of variables that researchers use to evaluate tactical behavior patterns: team balance, playing space, width, and length of playing space, and interpersonal distance. This categorization is a reference to frame the measures that researchers consider when analyzing the spatial behavior of players in both team and player levels.

Furthermore, another consideration when understanding spatial organization of players on the pitch is the team's formation. During a match, each team is lined up or arranged in a strategic formation on the pitch. Outfield players usually tend to encompass only a small portion of the pitch at any given instant. The team moves together as a cohesive unit to maintain their spatial arrangement on the pitch. Consequently, team formations are determined by the relative positions of the players. In other words, how players position themselves in relation to each other on the field defines the overall formation of the team during the game (Shaw & Glickman, 2019).

To analyze the spatial patterns created by players during a match, researchers have combined different geoprocessing tools as an aim to understand teams and players pattern behavior. In the next section, three geoprocessing tools are described to further analyze in the methodology section. Nearest distance calculation among pair of points is used to analyze an attacker-defender dyadic system and the relative distance to intercept a shot (Bauer et al., 2022; Vilar et al., 2012). Voronoi cells represent the dominant region of each player along the pitch which was also used to research around the distance between teammates (Fonseca et al., 2013). A convex hull represents the space covered by a team and can identify the phase a team is having either attacking or defending (Moura et al., 2012).

- *Nearest distance calculation*

According to Bauer et al. (2022), a defender can adopt the role of player-marking during an attack from the opposing team. In this specific role, a defender strategically positions themselves near their designated attacker, closely tracking their movements towards the target area. Ideally, the defender aims to maintain the closest possible distance to the attacker, typically positioning themselves on the side closer to their own goal.

Vilar et al. (2012) conducted a study focusing on attacker-defender dyadic systems, which occur when a 1 vs. 1 interaction takes place between two opponents during a game. The research aimed to explore how the positions of the goal and ball influence the pattern-forming dynamics of such dyadic systems. The findings indicated that both attackers and defenders tended to face the goal at similar angles, but defenders consistently remained closer to the goal than attackers. On the other hand, attackers consistently remained closer to the ball than defenders, while the defenders positioned themselves at a lower angle to the ball compared to attackers. At an individual level, the researchers observed that a stable state of coordination emerged when the attacker's distance to the ball was shorter than that of the defender, and the defender was positioned closer to, and between, the attacker and the goal.

Low et al. (2021) conducted an analysis of tactical behavior in football, specifically focusing on two pressing strategies: high-press defending and deep-defending. To gather data, the researchers used GPS devices on players during 72 trials of attack against defense in an 11 vs. 11 setup. They examined various measures, including the distance between teams and the distance to the nearest opponent (marking), among others. The study revealed some differences in marking strategies, but the authors emphasized that to understand patterns of tactical behavior, such as team formations, a more in-depth analysis of players' distances to their nearest opponents at a dyadic level is necessary.

- *Voronoi*

Voronoi diagrams are used to identify dominant regions on a football pitch. Fonseca et al. (2013) used Voronoi diagrams to research spatial dynamics of player's behavior in Futsal. They consider each Voronoi cell as the dominant region of a player. From a controlled experiment, they analyzed each player's Voronoi area and the nearest teammate distance at team level and individual level. They found that attackers have a larger dominant region compared to defenders. Furthermore, analyzing the nearest distance between teammates, they found that players from the attacker team tend to be further from each other in comparison with players from the defender team. This derives in the individual dominant region which was greater for the attacker team (team with the ball) than the defender team.

Caldeira et al. (2022) analyzed ten matches of the male French top football league (Ligue 1). They used Voronoi diagrams to understand the influence of team formation and player's roles within their dynamic interaction. The study revealed that team formations and player roles influence their connections with one another, leading to distinct spatial dominance patterns on the pitch. The researchers suggest that Voronoi diagrams can be transformed into meaningful compound variables, providing valuable insights into the matches. These insights can inform the development of representative training tasks to enhance the understanding of team dynamics and player interactions during gameplay.

– *Convex Hull*

Convex Hull is the total space covered by a team which is determined by calculating the boundary formed by the outermost players of the target team. (Coito et al., 2022). Moura et al. (2012) used Convex Hulls to calculate the area covered by teams and the dispersion of players on the field. The research revealed that during ball possession, players tended to be spread out over a larger area. Conversely, when the teams lacked ball possession, the players adopted a more compact formation, controlling a smaller pitch area. Shaw and Glickman (2019) used the Convex Hull as an exploratory tool observing that, although a team occupies various regions of the pitch at different moments, the players generally maintain their relative positions.

## 2.5. Conclusion

With the vast amount of spatiotemporal data available after a football match, the analysis of event data and trajectories has derived in the development of visual interfaces and visualizations that can provide insights or identify patterns, so to understand the game and provide meaningful performance information. It was possible to find certain parameters that football researchers use to evaluate tactical behavior at a player's and team's level (Coito et al., 2022).

To analyze the spatial patterns created by players during a match, researchers have combined different geoprocessing tools as an aim to understand teams and players pattern behavior. Three tools are mentioned in this theoretical background: nearest distance calculation, Voronoi diagrams and Convex Hull. Each of them has been used by researchers as tools to calculate or analyze measures derived from football datasets. In the methodology section, the three geoprocessing tools are used to output football geovisualizations to further evaluate them among users with and without previous football knowledge.

According to the literature review, table 1 shows the evaluation parameters selected to serve as an evaluation basis for football understanding. The parameters cover two levels of analysis: the team's and the player's levels. For a team's level analysis, researchers use the playing formation of a team, the dominant region of a team, and the playing space that a team covers throughout the match. For a player's level analysis, researchers consider the dyadic system of the attacker-defender distance, the relative distance for a defender to intercept a shot, and the distance between teammates. Regardless of the level of analysis, the parameters are linked and considered based on the geoprocessing tool analyzed.

| Table 1 Evaluation parameters | |
|---|---|
| **Parameters** | **Literature** |
| **Nearest Distance animation** | |
| Playing formation | Low et al. 2021 |
| Attacker-defender distance | Bauer et.al. 2022; Vilar et.al. 2012 |
| Relative distance to intercept a shot | |
| **Voronoi animation** | |
| Playing formation | Caldeira et al. 2022 |
| Dominant region | Fonseca et.al. 2013 |
| Distance between teammates | |
| **Convex Hull animation** | |
| Playing formation | Shaw and Glickman 2019 |
| Playing space | Moura 2012 |

# 3. Methodology

## 3.1. Introduction

This chapter will describe the methodology of the thesis into two parts. The first part focuses on the animations themselves. It starts describing the dataset I used for generating the animations and further describes the Python script used to apply the three geoprocessing tools described in the theoretical framework (Nearest distance calculation, Voronoi diagram and Convex Hull). The second part focuses on the evaluation of the animations generated, describing the survey methodology I used for the evaluation.

## 3.2. Animated Football Geovisualizations

- *Available Data*

I used an open dataset containing tracking data of a professional football match provided by Metrica Sports (Metrica Sports, n.d.). The company provides three datasets with tracking data from which I selected the first sample identified as Sample_Game_1. The data is in CSV format and is separated into two datasets: one for the home team and one for the away team. Both datasets have the same size of 145,006 records each, meaning that there is location in *x* and *y* of 22 players (11 for each team) and the ball for each frame (25 records per second). The position of the players and the ball goes from 0 to 1 in the 'X' and 'Y' axis. The cartesian coordinate system starts from the origin (0,0) in the top left corner until its limit at (1,1) coordinate in the bottom right. The field dimension considered for all the sample data is a standard 105 x 68 meters. The spatial resolution is 10 cm, and the temporal resolution is 25 frames per second. The sample is anonymized: players are identified with numbers and there are no details of names of either the teams or players.

- *Pre-processing Dataset*

I generated a Python script to access the data location in the web and used Pandas Python library to format and process the data. The format of the data was based on the requirements of the script that was used to generate the animations and that is described in the next pages. The script uses a Python DataFrame format as input data, and the structure of the headers is as follows:

["x", "y", "team_id", "player_id", "time"]

Table 2 describes the data types needed for the script to run. Based on the datatypes described, the dataset was adapted so to be able to run the geoprocessing tools script. It was necessary to rename the number of players so each of the teams had the same sequence from 1 to 11, starting from the goalkeeper and ending with the forwards. The dataset has information regarding the ID number of the frame and the time in seconds it refers to. For visualization reasons, the frame column performed better than the time in seconds.

| Table 2 Data types of the script inputs | | |
|---|---|---|
| **Column** | **Data type** | **Description** |
| 'x', 'y' | int/float | Player location coordinates in the 'x' and 'y' axis |
| team_id | int/string | Team Id for both attacking and defending teams |
| player_id | int/string | Player Id for both attacking and defending team. Id for ball is optional |
| time | int/float | Game time in seconds or any units |
| *Source: Kumar, S. 2019* | | |

To select the visually appropriate event of the game, I analyzed the dataset identifying key moments. Table 3 shows in detail an overview of the analysis I made for the game dataset and the key moments identified to generate the animations. After analyzing the dataset, I selected a subset of the data corresponding to 35 seconds of the game. The subset corresponds to a counterattack happening between seconds 1090 and 1125 (minute 18 of the game).

| Table 3 Analysis of the first 5 min of the dataset | | | | | |
|---|---|---|---|---|---|
| **Num.** | **Start time** | **End time** | **Frame starts** | **Frame ends** | **Details** |
| **1** | 0 | 30 | 0 | 749 | Start of the game. |
| **2** | 30 | 60 | 750 | 1499 | Corner for red team. |
| **3** | 60 | 90 | 1500 | 2249 | Arrangement of the teams. |
| **4** | 90 | 120 | 2250 | 2999 | First goal happens. |
| **5** | 120 | 150 | 3000 | 3749 | Celebration and restart of the game happens at s.147 |
| **6** | 150 | 180 | 3750 | 4499 | Ball posession on team blue and throw-in. |
| **7** | 180 | 210 | 4500 | 5249 | Change of possesion of the ball to team red. |
| **8** | 210 | 240 | 5250 | 5999 | Team red counterattack and attempt to goal. |
| **9** | 240 | 270 | 6000 | 6749 | Goal kick of team blue. |
| **10** | 270 | 300 | 6750 | 7499 | Attempt to goal from team blue. |
| *Source: Sample Dataset 1* (Metrica Sports, n.d.) | | | | | |

- *Script*

The script used for creating the football animations was created by Samira Kumar (2019) and is based in the Bokeh library. The script performs better in a Jupyter Notebook environment. It consists of three Python files which call the functions for visualizing first the Voronoi diagram and Convex Hull together, and the Nearest Distance calculation is in a separate file.

- *Parameters*

The scripts consider twelve input parameters to run. The user defines the IDs for the attacking and defending team according to the dataset, as well as the range or the coordinate system of the location data. It is possible to add a background image for the visualization. The animation speed and the number of frames seen can be modified. The script is suited for visualizing different sports. Table 4 gives an overview of all the parameters needed for the script to run.

**Table 4 Input parameters for geoprocessing tool script**

| Parameter | Description |
|---|---|
| doc | Plots the graph |
| df | Gets the user defined dataframe |
| headers | Give the headers to the dataframe |
| id_def | Provide id of defending team |
| id_att | Provide id of attacking team |
| x_range | Provide 'x' range of the pitch coordinates |
| y_range | Provide 'y' range of the pitch coordinates |
| image_url | Provide the location of the background image of the pitch |
| slider_steps | Provide the slider steps (number of slides to show by the speed defined) |
| sport | Provide the sport details to change slider function |
| anim_speed | Provide speed of animation in milliseconds |
| show_dist_speed | Turn on/off plotting speed and distance |
| *Source: Kumar, S. 2019* | |

For the dataset used, the IDs of the teams were 1 for the attacking team and 2 for the defending team, nevertheless, to better understand the visualization, I refer to the teams as 'red team' or the attacking team and 'blue team' as the defending team. I assigned the ball the ID 3. MetricaSports uses a coordinate system that goes from the origin (0, 0) until its limit (1, 1). The script also considers when the location coordinates are beyond limits, e.g., when players go beyond the borders of the pitch. For the timeframe selected (35 seconds) the tool displayed 874 frames. The time resultant was 25 seconds of animation.

- *Modularization*

The next table describes the Python script used for generating the animations. The script returns an interactive animation which can be manipulated by the user to display or not the Voronoi diagram or the Convex Hull. For the Nearest Distance calculation animation, a different script is used, although, the structure is the same but the module *cdist* from *scipy* is used for the distance calculation.

| Table 5 Modularization of the Python script used | | |
|---|---|---|
| **Function** | **Parameters** | **Description** |
| def make_plot | (doc, df, headers, id_def, id_att, slider_steps, x_range, y_range, image_url, sport='football', anim_speed=50, show_dist_speed=False) | Input parameters |
| def patches_from_voronoi(vor): | vor - Output from the scipy Voronoi | Function runs in a separate script. Returns the x and y values to plot patches and boundary lines |
| def get_convex_hull | (team_def, team_att, current_time) | Create the convex hull for the coordinates |
| def plot_clean | (plot) | Remove plot background and alter other styles |
| def update_data | (attrname, old, new) | Update the figure every time slider is updated |
| def animate_update | Animation | Allows to modify the slider steps for the animation |
| def animate | Animation | Defines buttons: play and pause |
| *Source: Kumar, S. 2019* | | |

- *Outputs*

The IDE (Integrated Development Environment) I used to retrieve the data and generate the output animations was Jupyter Notebook. As a first step, I called the data and modified it according to the requirements of the geoprocessing tool to run. Then, I used a GIF generator to record the animations in .gif format. Four animations were created, one raw animation (Figure 1), meaning without geoprocessing tools. The other three are described below.

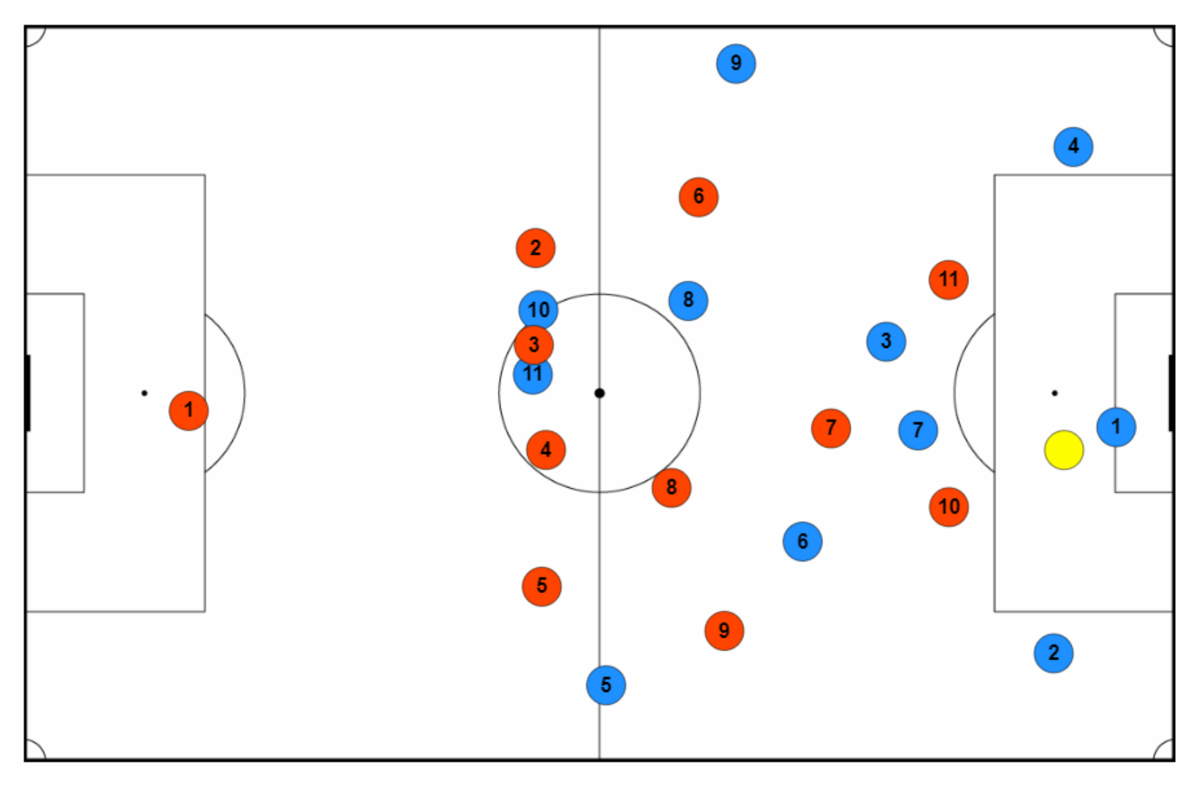


*Figure 1 Raw animation. Red team (attacking team) and Blue team (defending team) are displayed. The ball is in yellow color.*

- *Nearest distance player*

The tool considers the nearest distance from each of the defending team players to the attacking team players and displays a dotted line from the defending team player to the nearest attacking player, defending players can have more than one line or more than one closer attacker. In the event considered, the red team starts as the defending team, as the blue team has the possession of the ball. Once the blue team loses the ball, the distance calculated changes to the opposite, the blue team is defending and the red team attacking. Figure 2 shows the resultant animation.

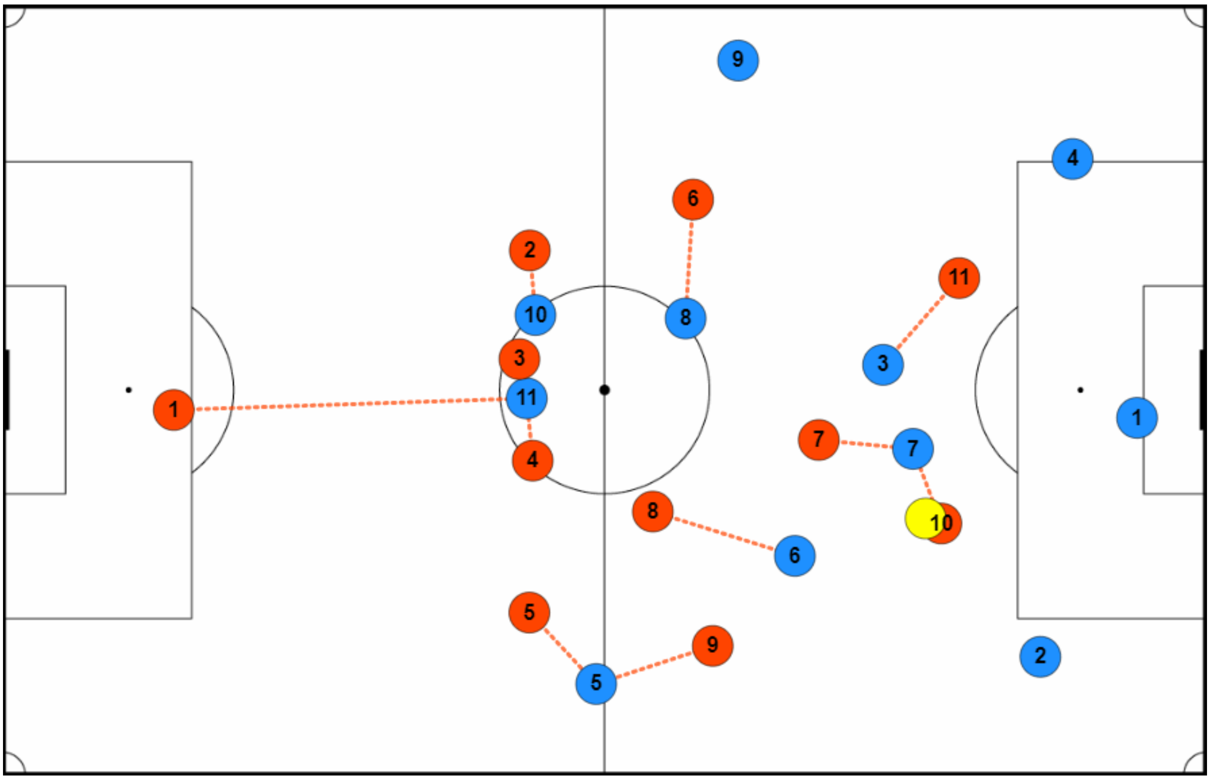


*Figure 2 Nearest Distance calculation. A dotted line displays showing the nearest distance attacker to a defender player.*

- *Voronoi*

The Voronoi diagram generates cells that are considered dominant regions for each player. The ball is also considered for the calculation of the Voronoi. According to the spatial distribution of the players, the tool updates each position of the players for the generation of the Voronoi diagrams. The lines of the diagram are of dark red color to avoid a distraction of the user and the fill color of the cells is avoided to direct the attention of the viewer to the entire animation and not to specific cells. Figure 3 shows the resultant Voronoi animation.

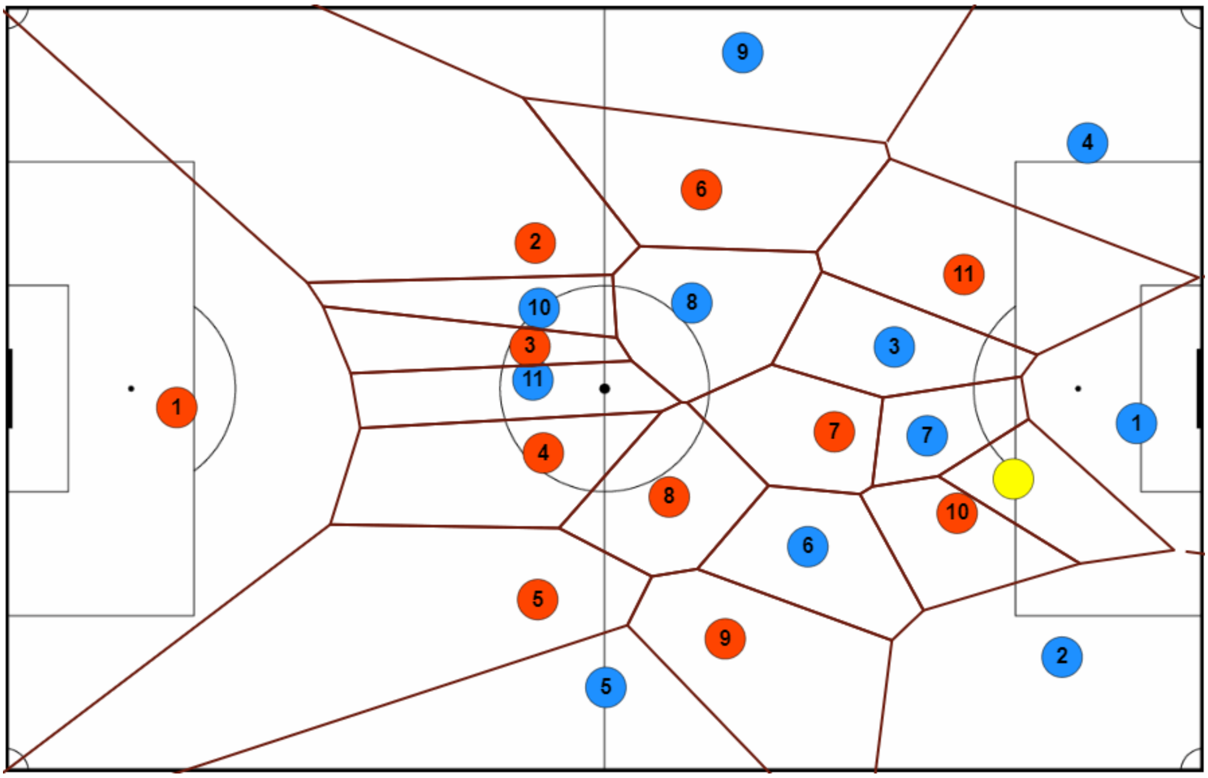

*Figure 3 Voronoi diagram generated considering the position of the players of both teams and the ball to generate the Voronoi cells.*

- *Convex Hull*

Figure 4 shows the resultant animation displaying two Convex Hulls for the two teams, red and blue. The polygons are constructed based on the outer players of each team; the position of the ball is avoided for this animation. The polygons have the same color of the teams and when overlayed each other, they have a transparency for visualization purposes.

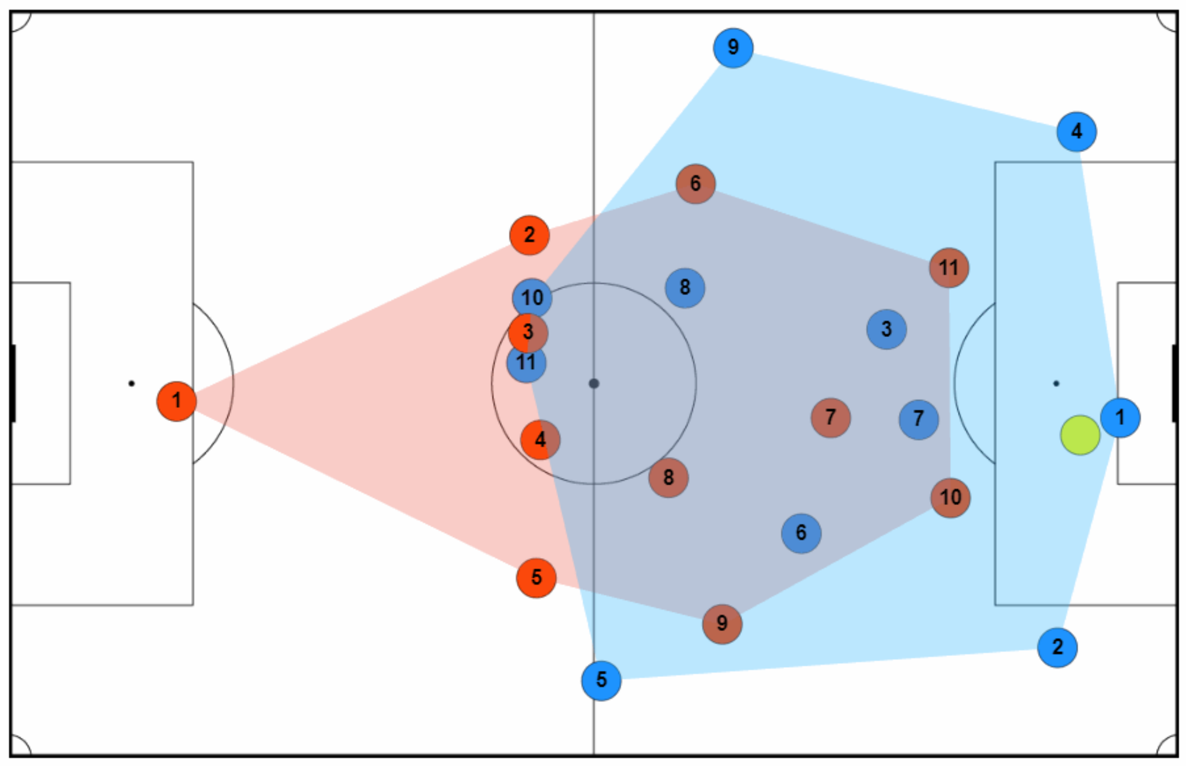


*Figure 4 Convex Hull displayed for each team.*

## 3.3. Evaluation

The animations are evaluated based on a questionnaire that is part of a survey. I used ArcGIS Survey123 (https://survey123.arcgis.com/) from ESRI to design and conduct the survey. This tool provides an online builder and a desktop version. For design purposes, I first used the online version, which has an intuitive user interface and makes it easier to create the overall survey and modify aspects of the design such as colors. After a draft of the design, I used the desktop version that is based in an XML file that lists questions and answers accordingly. The XML version allows to upload images with sizes over 1MB, therefore, it was necessary to use the advanced tool, since the animations average size is 7MB.

- *Survey structure*
  - *Information and Informed Consent*

The survey starts with detailed information about the research and purpose of the survey. It includes contact details so the participant can ask further questions about the research. The informed consent informs the participant about the use of the data gathered and possible research results publication, so the data can be treated accordingly.

– *Demographic information*

Two parameters regarding demographics are considered for the survey: gender and sex. These parameters are considered to analyze the results based on demographics in contrast to football knowledge.

– *Football Knowledge*

The survey includes a section to characterize and categorize the participants in their football knowledge. Three questions are displayed on one page going from easy to difficult.

– *Training*

A training section follows to familiarize the participant with the shape, speed, and type of animation they will visualize during the evaluation. The training section includes basic information about the sport and mentions all the terms used in the evaluation questions to help the participant to not get lost during the evaluation. One animation from a different event of the game is added and three static images.

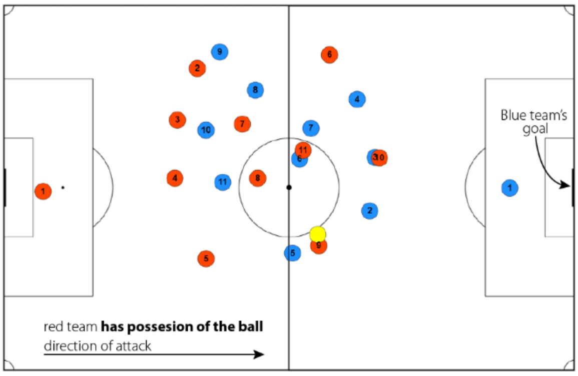


*Figure 5 Training Information about red team.*

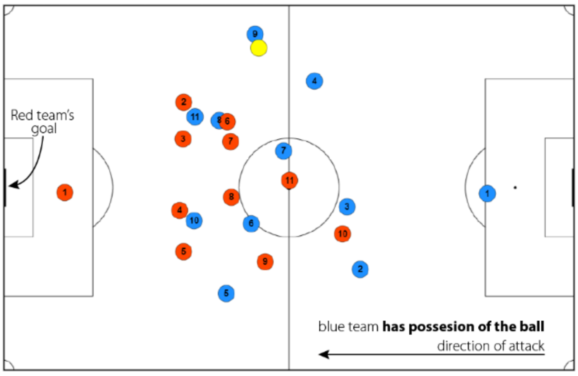


*Figure 6 Training information about blue team*

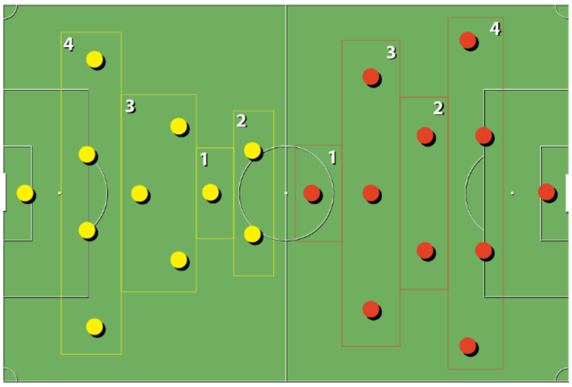

*Figure 7 Formation example*

– *Evaluation*

The evaluation section displays each animation and the corresponding questions in a single page of the survey. The answers are of multiple choice.

- *Participants*

The target participants are users with football knowledge and without football knowledge. The football section was added to filter the users according to their football knowledge. It has three questions that are used to categorize the type of participants and they have the possibility to make subcategories. Table 5 shows the categorization of the participants according to the questions.

| ***Table 6 Participant categories*** | | | |
|---|---|---|---|
| **Section: Football knowledge** | | | |
| **Num.** | **Question** | **Category** | **Sub-category** |
| | First wrong answer) | Zero knowledge group | Zero knowledge group |
| **Q1** | How long does a football match take? | Knowledge group | Basic knowledge |
| **Q2** | Which team won the last FIFA World Cup? | | Medium knowledge |
| **Q3** | From the image above, what is the lineup of the teams? | | Advanced knowledge |

- *Questionnaire design*

The evaluation section consists of six pages that display raw animations and nine questions asked for the evaluation. The animations are presented in a pseudo-randomized order to try to avoid the learning effect for the participant. Each of the three types of geovisualizations has its own raw image which intends to be compared with.

| **Table 7 Order of animations** | |
|---|---|
| Animation 1 | raw |
| Animation 2 | voronoi |
| Animation 3 | raw |
| Animation 4 | marking |
| Animation 5 | raw |
| Animation 6 | convex hull |

The questions are designed based on specific measures found in literature that are described in the theoretical background chapter. These measures are used to analyze spatial patterns based on geoprocessing tools, such as the evaluated in this thesis: nearest distance calculation, Voronoi diagrams and convex hull. Table 7 cites each question with the parameter used in the corresponding literature.

| Table 8 Evaluation Parameters | | | |
|---|---|---|---|
| **Num.** | **Question** | **Parameter** | **Literature** |
| **Marking player animation** | | | |
| 1 | What is the lineup or formation of the red team? | Playing formation | Low et al. 2021 |
| 2 | When the blue team has possession of the ball, how would you describe the positioning of the defenders from the red team? | Attacker-defender distance | Bauer et.al. 2022; Vilar et.al. 2012 |
| 3 | What was the role of player #2 in the red team? | Relative distance to intercept a shot | |
| 4 | Ideally, a defender should be positioned between the goal and the attacker. Based on your observations, the player #2 from the red team is: | Relative distance to intercept a shot | |
| **Voronoi animation** | | | |
| 5 | What is the lineup or formation of the red team? | Playing formation | Caldeira et al. 2022 |
| 6 | When does a team cover a larger area of the pitch? | Dominant region | Fonseca et.al. 2013 |
| 7 | In what situations do a team's players get close to each other? | Distance between teammates | |
| **Convex Hull animation** | | | |
| 8 | What is the lineup or formation of the red team? | Playing formation | Shaw and Glickman 2019 |
| 9 | When does a team cover a larger area of the pitch? | Playing space | Moura 2012 |

- *Procedure*
  - *Pilot tests*

I conducted two pilot tests to evaluate three aspects: methodology of evaluation, grammar and clearness, and aesthetics.

*First pilot test*

The first pilot test was conducted among my thesis supervisor, a writing coach (Writing Center TUDresden), and a cartography colleague. The design of the evaluation questionnaire was the most criticized and got the most feedback. The goal was to link some literature regarding spatial point pattern analysis with geoprocessing tools, with the animations and elaborate a question that can link both. Another feedback comment was to include a training section so to give a context to the participant before the evaluation section.

The grammar and clearness of the survey was successfully edited based on the feedback from the writing coach of the writing center of TUDresden. The feedback on the aesthetics of the animations from my cartographer colleague was useful to maintain the same line of design among all the visualizations. Redundance, clearness and grammar feedback were collected among the three participants.

### *Second pilot test*

I conducted a second pilot test to get feedback in a systematic way. The same parameters to evaluate were lack of context, clearness, wording, and aesthetics of the animations. The survey was the draft updated with the feedback from the first pilot test, but with additional questions divided into two sections. First, a Likert scale to evaluate the parameters mentioned above; and second, a list with the nine evaluation questions with an empty free text box, so each participant can leave feedback for each specific question. Seven people answered the questionnaire. All the respondents were cartography and geomatic master students. Table 8 shows the results of the Likert scale for each of the statements considered.

**Table 9 Second pilot test results**

| Parameter | Num. | Statement | Strongly disagree | Disagree | Neutral | Agree | Strongly agree |
|---|---|---|---|---|---|---|---|
| Lack of context | 1 | The training information provided sufficient context for me to understand the tasks to solve. | | | 1 | 4 | 2 |
| Clearness | 2 | I found all the questions and tasks easy to comprehend. | | | 1 | 3 | 3 |
| | 3 | The statements of the questions and tasks were unambiguous. | | 2 | 2 | 2 | 1 |
| Wording | 4 | The training section included all the football terms used in the questionnaire. | | | 2 | 3 | 2 |
| Animation design | 5 | I can clearly differentiate between the teams through the selected colors for the animation. | | | | 2 | 5 |
| | 6 | I was able to easily identify the ball in the animation. | | | | 1 | 5 |
| | 7 | The animation speed was appropriate, allowing me to answer the questions and solve the tasks effectively. | | | | 4 | 3 |

Some feedback was collected for each of the statements as it follows:

> **Statement:** *The training information provided sufficient context for me to understand the tasks to solve.*
> **Feedback:** “Maybe it would've been nice to see the formation of the team at a point midgame to clarify which formation (4-4-2, ...) is shown there.”
> **Question:** *What is the lineup or formation of the red team?*
> **Feedback:** “maybe just use either lineup or formation and not both :)”
> **Question:** *What was the role of player #2 in the red team?*
> **Feedback:** “There is a typo in the answers, should be role instead of roll.”

I updated the survey based on the feedback collected and made it publicly available.

- *Data collection*

The survey was open from July 28 until August 30, 2023. Participants were recruited mostly via word of mouth and via flyers posted around the Cartography department of TUMunich. The flyer had some explanation about the research and the QR and link to the survey. The ArcGIS123 tool used for designing and conducting the survey provides the possibility of filling out it on different electronic devices (website, smartphone, tablet).

# 4. Results and discussion

## 4.1. Introduction

Results are presented based on the answers to the questions designed to test the three geoprocessing tools evaluated in this research in the same order: Nearest Distance calculation, Voronoi diagrams, and Convex Hull. After presenting the results, each section discusses the results, showing a final graph summarizing the findings. At the end of the chapter, the discussion aims to answer the research question.

The survey started with questions aimed at characterizing the participants regarding age and sex and followed a football knowledge section that was used to compare the difference between participants who know about football and those with little knowledge. Next, the evaluation section started with the questionnaire that evaluates the performance of the animations. Geovisualization is the term for referring to animations that have a geoprocessing tool in them. The logic of the survey was to compare each geoprocessing tool animation with a plain or raw animation (meaning that it shows only points as the location of players and the ball), see Figure 8.

The survey collected data from 109 participants, of which 57 were male and 48 female. The remaining identified themselves as non-binary or did not want to say. All the participants filled out the survey entirely. The average age of most participants (76%) ranges between 21 and 30. Participants with football knowledge represent 64% of the sample (n=70). The condition to categorize a participant with football knowledge was to correctly answer all the questions from the *football knowledge* section. From this group, 37% of the sample (n=26) are female and 60% (n=42) males. The remaining 3% (n=2) preferred not to identify their gender. The remaining 36% (n=39) of the sample represents participants without football knowledge. From this group, 56% (n=22) are female and 39% (n=15) are males. The remaining 5% (n=2) are non-binary.

The description of the results starts with a separate analysis of the first question about the formation of a team because the question is the same for all the animations. Next, the results are described according to each Geoprocessing tool analyzed, starting with the Nearest Distance calculation, the Voronoi diagrams, and the Convex Hull. Table 10 shows the parameters analyzed through each question and table 11 presents the results for each of the option considered (correct, incorrect and not answered question) for the total sample and the two groups analyzed.

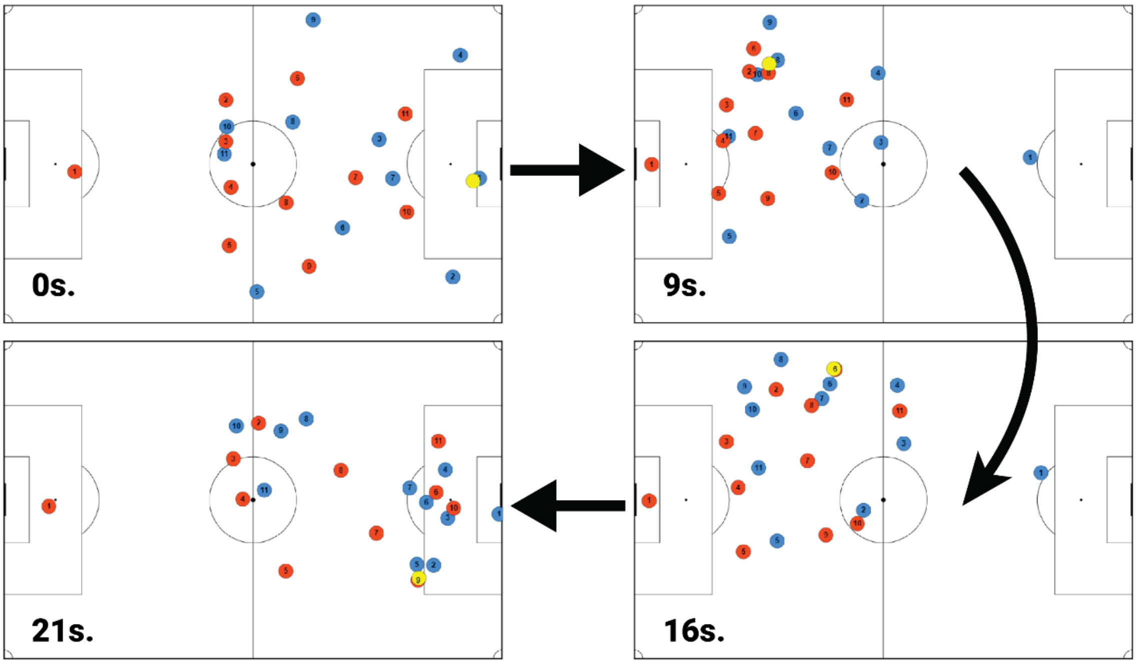


*Figure 8 Four static images of the Raw animation or without a geoprocessing tool.*

| Table 10 Questionnaire | |
|---|---|
| **Parameter** | **Question** |
| **Point patterns** | |
| Playing formation | What is the lineup or formation of the red team? |
| **Nearest distance** | |
| Attacker-defender distance | When the blue team has possession of the ball, how would you describe the positioning of the defenders from the red team? |
| Relative distance to intercept a shot | What was the role of player #2 in the red team? |
| | Ideally, a defender should be positioned between the goal and the attacker. Based on your observations, the player #2 from the red team is: |
| **Voronoi diagrams** | |
| Dominant region | When does a team cover a larger area of the pitch? |
| Distance between teammates | In what situations do a team's players get close to each other? |
| **Convex Hull** | |
| Playing space | When does a team cover a larger area of the pitch? |

**Table 11 Questionnaire results for the total sample and the two groups analyzed**

| Geoprocessing tool | Football parameter | Total sample (n=109) | | | | | | Participants With football knowledge (n=70) | | | | | | Participants Without football knowledge (n=39) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Raw Animation | | | Geoprocessing tool | | | Raw Animation | | | Geoprocessing tool | | | Raw Animation | | | Geoprocessing tool | | |
| | | Correct | Incorrect | Not Answered | Correct | Incorrect | Not Answered | Correct | Incorrect | Not Answered | Correct | Incorrect | Not Answered | Correct | Incorrect | Not Answered | Correct | Incorrect | Not Answered |
| Nearest Distance | Playing formation | 76 | 21 | 12 | 79 | 17 | 13 | 55 | 11 | 4 | 57 | 10 | 3 | 21 | 10 | 8 | 22 | 7 | 10 |
| | Attacker-defender distance | 60 | 40 | 9 | 68 | 38 | 3 | 36 | 29 | 5 | 40 | 30 | 0 | 24 | 11 | 4 | 28 | 8 | 3 |
| | Relative distance to intercept a shot | 101 | 7 | 1 | 96 | 9 | 4 | 69 | 1 | 0 | 65 | 3 | 2 | 32 | 6 | 1 | 31 | 6 | 2 |
| | | 99 | 5 | 5 | 94 | 4 | 11 | 65 | 3 | 2 | 62 | 4 | 4 | 34 | 2 | 3 | 32 | 0 | 7 |
| Voronoi | Playing formation | 91 | 7 | 11 | 94 | 1 | 14 | 63 | 3 | 4 | 67 | 0 | 3 | 28 | 4 | 7 | 27 | 1 | 11 |
| | Dominant region | 91 | 13 | 5 | 92 | 10 | 7 | 58 | 10 | 2 | 62 | 5 | 3 | 33 | 3 | 3 | 30 | 5 | 4 |
| | Distance between teammates | 90 | 14 | 5 | 89 | 15 | 5 | 58 | 10 | 2 | 59 | 10 | 1 | 32 | 4 | 3 | 30 | 5 | 4 |
| Convex Hull | Playing formation | 85 | 15 | 9 | 85 | 14 | 10 | 63 | 6 | 1 | 59 | 8 | 3 | 22 | 9 | 8 | 26 | 6 | 7 |
| | Playing space | 95 | 10 | 4 | 98 | 7 | 4 | 62 | 6 | 2 | 63 | 6 | 1 | 33 | 4 | 2 | 35 | 1 | 3 |

## 4.2. Evaluation

- *Point patterns - team formation*

**Question**: *What is the lineup or formation of the red team?*

One of the basic principles of understanding football tactics is to identify a team's formation on the field. Researchers use geoprocessing tools to identify team formation that can vary along the match. Depending on the focus of the analysis, the researchers use a specific tool to visualize the formation of a team. Among the three geoprocessing tools analyzed in this research, the literature implies that identifying a team formation is a prior step for in-depth analysis at the individual or player's level (Caldeira et al., 2022; Low et al., 2021; Shaw & Glickman, 2019). Therefore, the survey aimed to compare the understanding of team formation among the participants with the four types of animation: raw, Nearest Distance, Voronoi, and Convex Hull.

The whole sample shows different results and not a similar trend among the compared animations. Correct answers were more for the Nearest Distance (79) and the Voronoi animation (94). The Convex Hull animation registered the same number of correct answers (85) for both the raw animation and the one that has the geoprocessing tools included. For the Nearest Distance and Convex Hull animations, nearly 15% of the sample replied incorrectly to the question, and 10% chose not to answer the question. The geoprocessing tool animations have more non-answered replies than the raw animation, with a difference of about 1 point.

– *Participants with football knowledge*

In the case of the group of participants with football knowledge, the Voronoi animation had an increase in correct answers by 6 points and 0 incorrect answers. The Nearest Distance animation also increased by 2 points, although incorrect answers remain around 15%. In contrast, the Convex Hull animation decreased correct answers by 4 points.

– *Participants without football knowledge*

The participants without football knowledge show an increase in correct answers comparing the raw animation with the geoprocessing tool animation for the Nearest Distance calculation and the Convex Hull animation and a decrease in correct answers for the Voronoi animation. The Voronoi animation decreased by 2 points of correct answers and an increase of participants that preferred not to answer the question by 10%.

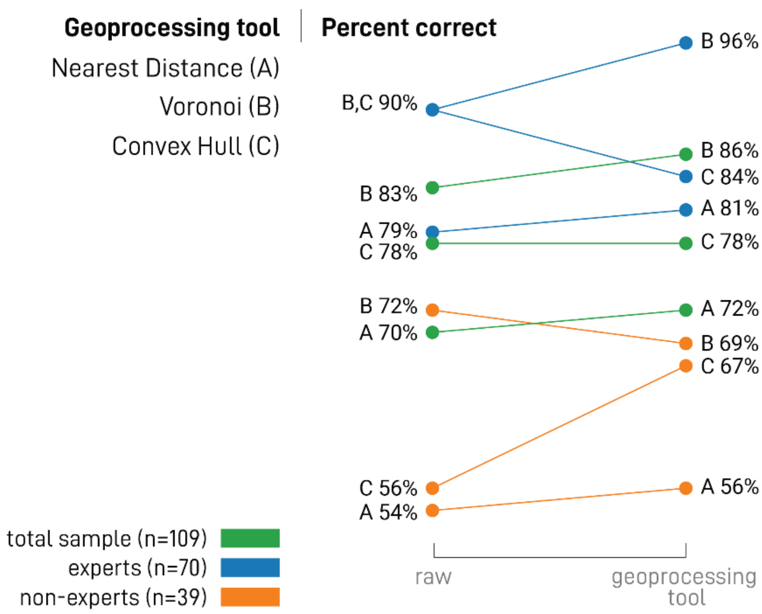


*Figure 9 Percentage of correct responses for the team formation parameter.*

When comparing the results of the participants with football knowledge with the ones without football knowledge, it is possible to observe differences and not a specific trend following both groups. In the case of the Nearest Distance animation, both groups had a slight increase in correct answers. However, the Geoprocessing tool promoted a reluctance to answer the question for the participants without football knowledge. For the Voronoi diagram, there are differences between both groups. The users with football knowledge improved the number of correct answers and did not register incorrect answers for the Voronoi animation. On the contrary, users without football knowledge showed decreased correct answers when looking at the animation with the geoprocessing tool. According to the results, the Convex Hull animation helped the users without football knowledge to answer the question correctly. It encouraged participants to answer the question and not avoid it. In contrast, participants with football knowledge show that the geoprocessing tool did not help on improving the correctness of the answers.

- *Nearest distance calculation*

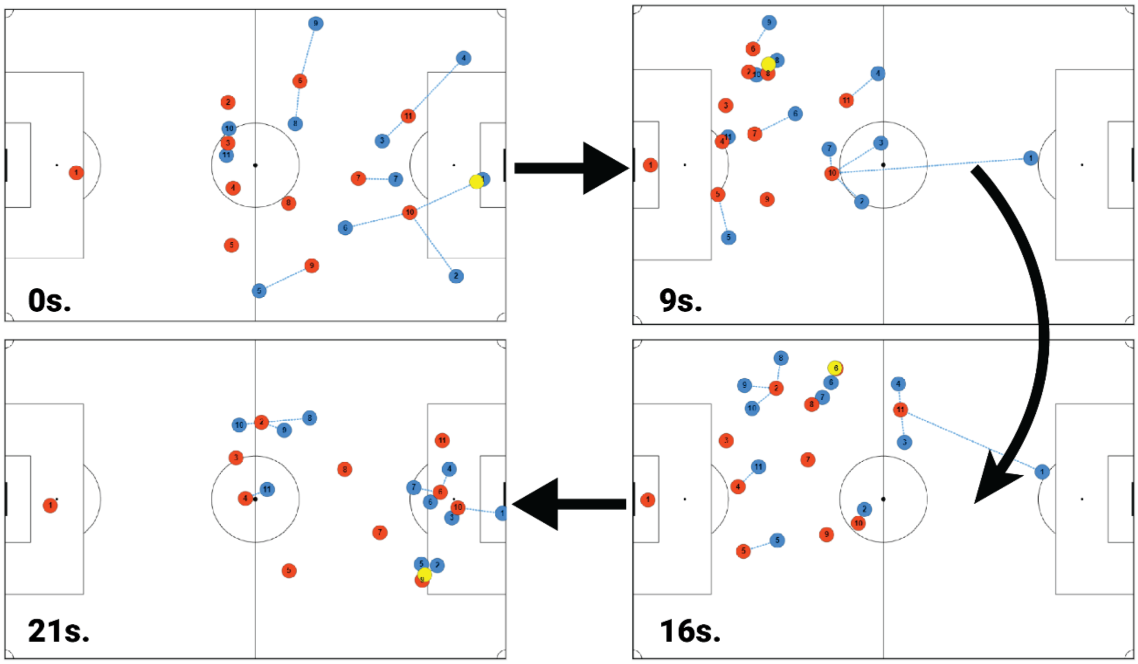


*Figure 10 Four static images of the animation using the Nearest Distance calculation tool.*

**Parameter: Attacker-defender distance**

**Question:** *When the blue team has possession of the ball, how would you describe the positioning of the defenders from the red team?*

In dyadic systems like an attacker-defender situation in a football game, the location of the ball does not influence the coordinated position of each of the players as the location of the goal (Vilar et al., 2012). With this statement in mind, the question aimed to understand and observe the coordination of the attacker-defender dyadic system in a team. Results show that the Nearest Distance animation helped increase the total sample's understanding by 7 points. The number of incorrect answers decreased from 40 for the raw animation to 38 for the geovisualization, and it also motivated the participants to answer the question, observed in the reduction of the non-answered questions by 5 points.

**Parameter: Relative distance to intercept a shot**

**Question 1:** *What was the role of player #2 in the red team?*

At a player level of analysis, in an attacker-defender situation, the distance from the attacker to the ball is less than the distance from the defender to the ball. The defender's position should also be between the ball and the defended goal and much closer to the goal (Vilar et al., 2012). I used two questions to evaluate the understanding of the principle of the relative distance to intercept a shot based on the role of a defender.

For the first question, regarding the role of defender #2 of the red team, all the participants performed better with the raw animation, having 101 correct answers compared to the Nearest

Distance animation (96). The number of incorrect answers increased slightly from 7 for the raw animation to 9 for the Nearest Distance animation. Non-answered questions increased from 1 for the raw animation to 4 for the Nearest Distance animation.

**Question 2:** *Ideally, a defender should be positioned between the goal and the attacker. Based on your observations, the player #2 from the red team is:*

The second question regarding the position of the defender of the red team showed fewer correct answers when using the Nearest Distance animation. Correct answers for the raw animation were 99 compared to the 94 received in the geovisualization animation. The raw animation had five responses for the incorrect and the same number for the non-answered questions. The Nearest Distance animation had four incorrect answers (less than the raw animation). However, more participants (11) chose not to answer the question.

- *Participants with football knowledge*

**Parameter: Attacker-defender distance**

The Nearest Distance animation performed better than the raw animation. The correct answers increased by 6 points. All participants answered correctly or incorrectly with the Nearest Distance animation; therefore, the option to not answer the question got 0 responses.

**Parameter: Relative distance to intercept a shot**

Like the overall trend, participants with football knowledge show fewer correct answers for the animation using the Nearest Distance calculation. Nevertheless, the number of correct answers decreased by 6 points in the first question of this group and by 4 points in the second question. Incorrect answers increased, confirming that participants performed better with the raw animation. The Nearest Distance animation made 2 participants decide not to answer the question in both cases.

- *Participants without football knowledge*

**Parameter: Attacker-defender distance**

The users without football knowledge have the same trend as the overall sample and the participants with football knowledge. The Nearest Distance animation helped to understand better the attacker-defender distance principle. For the users without football knowledge, the Nearest Distance animation had more correct answers (28) than the raw animation (24), showing an increase of 10%. The incorrect and not answered responses decreased inversely proportional to the correct ones.

**Parameter: Relative distance to intercept a shot**

In the case of the participants without football knowledge, the trend remained with a decrease in the number of correct answers from the Nearest Distance animation compared to the raw animation. The raw animation registered more correct answers (32) than the Nearest Distance animation (31) for the first question. The second question got 34 correct answers for the first question and 32 correct answers for the geovisualization animation. The incorrect answers were 6 for both cases. The second question got 0 incorrect answers for the Nearest Distance animation, although the number of participants who did not answer increased from 3 for the raw animation to 7 for the geovisualization animation.

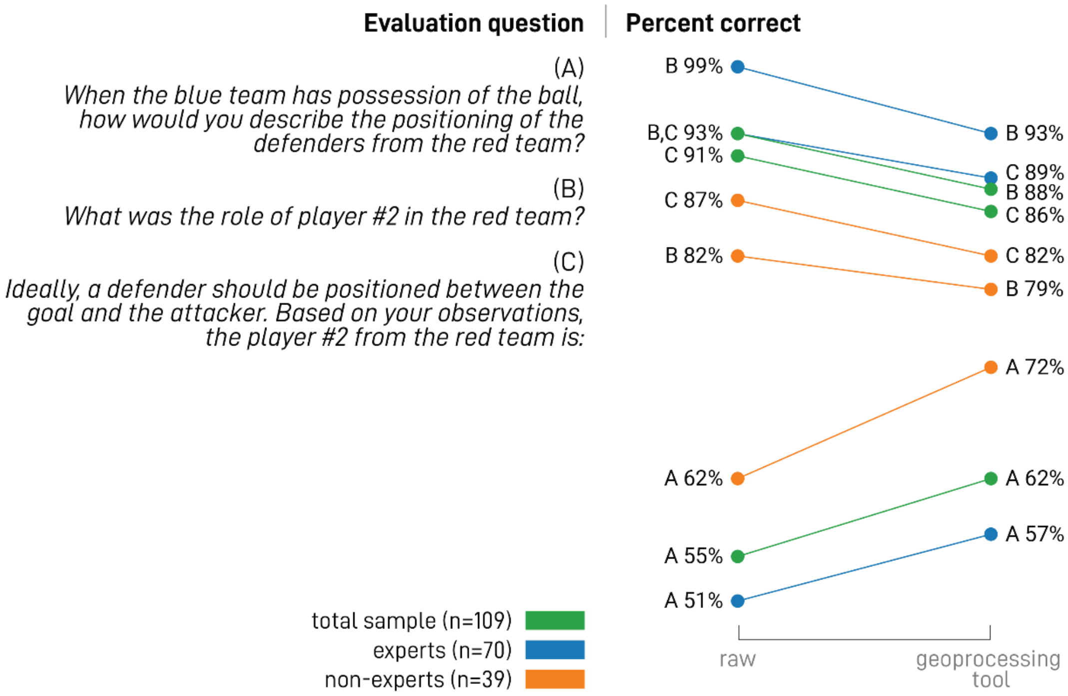


*Figure 11 Percentage of correct responses for the Nearest Distance calculation animation.*

The emphasis on visualizing the Nearest Distance between players aims to help the user visualize the dynamic relationship between the attacker and defender team at a player level. The tool calculates the Nearest Distance between the players from the defending team and the closest attacker. The animation shows a dotted line that establishes the relationship between players. The color of the line corresponds to the defending team; therefore, at the second 12 in the animation, the role of the blue team changes from attacker to defender. The color of the distance line turns blue.

From the results, both groups, participants with and without football knowledge, show that the geoprocessing tool animation got more correct answers only for the first question (*When the blue team has possession of the ball, how would you describe the positioning of the defenders from the red team?*). Interestingly, participants with football knowledge showed an unusual number of participants that answered incorrectly, nearly 40% of the sample (n=70). The reason can rely on a football background knowledge expertise with a critical point of view that considers details that can go beyond the aim of the question.

- *Voronoi diagrams*

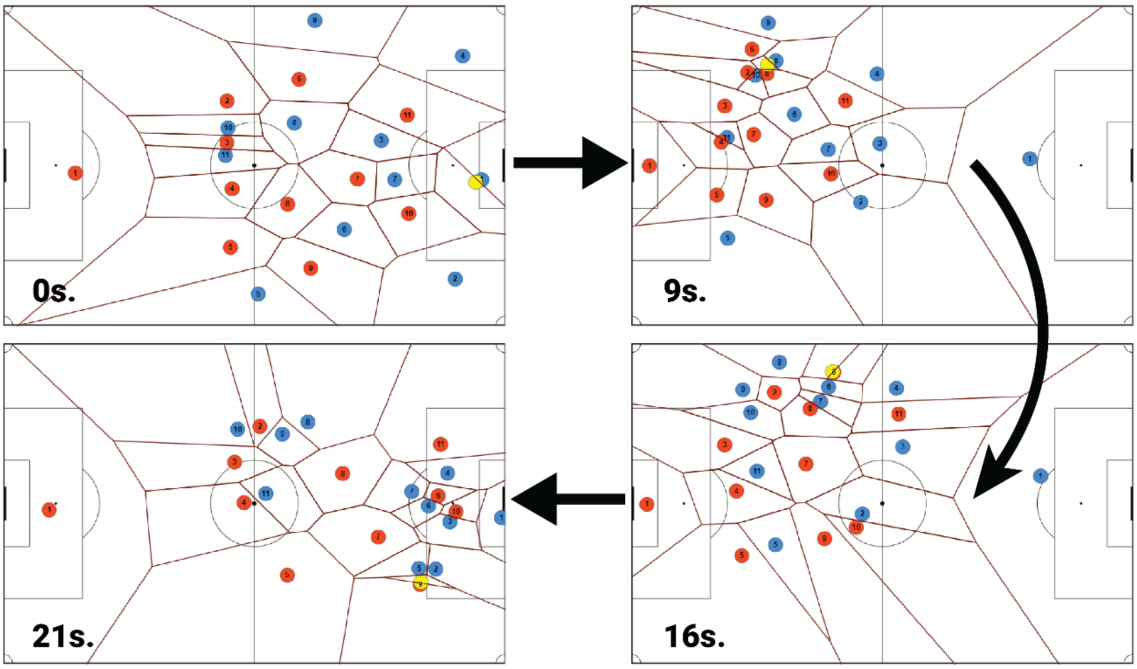


*Figure 12 Four static images of the animation using the Voronoi diagram tool.*

**Parameter: Dominant region**

**Question:** *When does a team cover a larger area of the pitch?*

Researchers use Voronoi diagrams to analyze the dominant region of players. When a team is attacking, every player tends to have a bigger Voronoi cell on average. Therefore, the attacking team has a greater dominant region than the defending team (Fonseca et al., 2013). The whole sample shows a slight improvement in understanding the dominant region covered by players of the attacking team, showing 92 correct answers for the geovisualization compared to the 91 that the raw animation got. The incorrect answers for the Voronoi animation were less (10) than for the raw animation (13). In contrast, more participants preferred not to answer the question for the geovisualization animation (7) compared to the 5 participants who did not want to answer for the raw animation.

**Parameter: Distance between teammates**

**Question:** *In what situations do a team's players get close to each other?*

At a player-level analysis, when a team is attacking, the distance among the teammates tends to be further from each other than the defender team (Fonseca et al., 2013). Contrary to the first question, participants showed a slightly worse understanding of the distance between teammates principle. The raw animation had 90 correct answers compared to 89 for the geovisualization. Incorrect answers increased from 14 for the raw animation to 15 for the geovisualization. The number of participants remained the same (5) for both types of animation.

– *Participants with football knowledge*

Users with football knowledge showed an increase in the correct answers in both parameters, the dominant region and the distance between teammates. The raw animation got 58 correct answers in both cases; the Voronoi diagram animation increased the number of correct answers for the first question by 6 (62) points and for the second question by 1 point (59). For the first question, the incorrect answers decreased from 10 for the raw animation to 5 for the geovisualization animation. For the second question, the incorrect answers remained the same for both cases (10 responses). The participants chose to answer more about the Voronoi animation than the raw animation in the second question.

– *Participants without football knowledge*

In contrast to the participants with football knowledge, users without football knowledge had a decrease in correct answers for both questions. The raw animation got 33 and 32 correct answers, respectively, compared to the 30 responses that the Voronoi animation got for both questions. On the contrary, the incorrect answers increased for both cases, getting five responses for the Voronoi diagram. The participants that chose not to answer were 3 for the raw animation and 4 for the geovisualization in both cases.

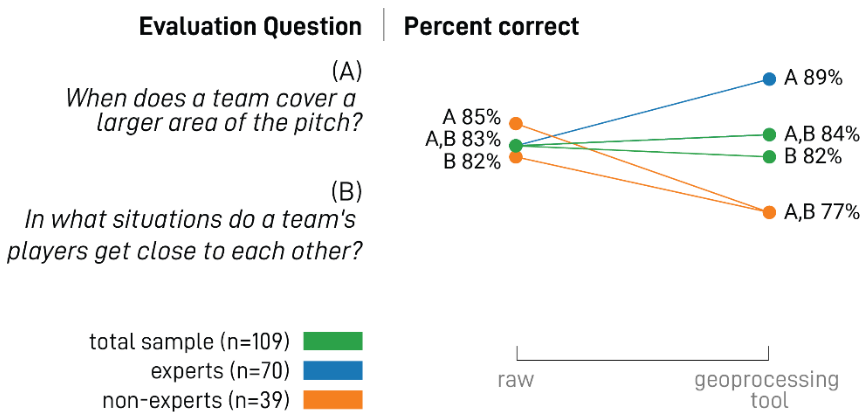


*Figure 13 Percentage of correct responses for the Voronoi diagram animation.*

This geoprocessing tool helped the users with football knowledge more than the participants without football knowledge. However, the increment of correct answers is not more than 1% of the sample (n=70). In contrast, the participants without football knowledge showed that the Voronoi diagrams did not help visualize and understand the dominant region of the players, nor the distance between the teammates. As with the participants with football knowledge, the difference with the correct answers for the raw animation is not more than 1%.

- *Convex Hull*

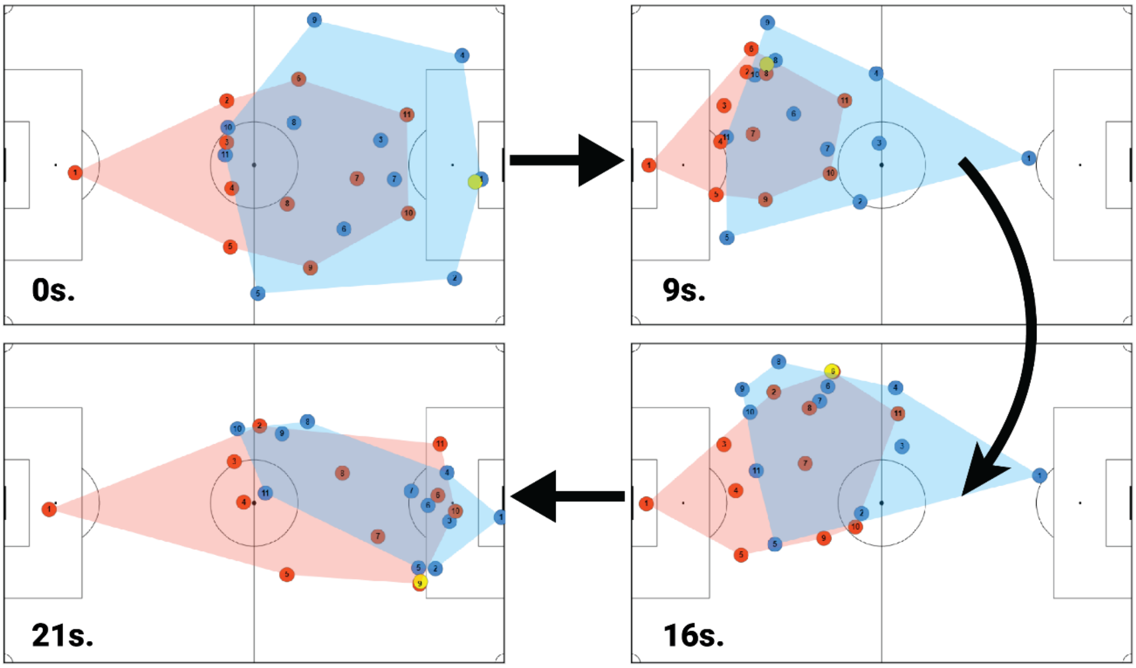


*Figure 14 Four static images of the animation using the Convex Hull tool.*

When a team is attacking or has ball possession, the players disperse themselves to cover a greater area on the pitch (Moura et al., 2012). The participants showed a better understanding of the playing space of a team when seeing a geovisualization. Correct answers for the Convex Hull animation were 98 compared to the 95 correct answers for the raw animation. The understanding is also evident in the decrease of the incorrect answers from 10 for the raw animation to 7 for the geovisualization. The participants who chose not to answer the question remained the same (4).

– *Participants with football knowledge*

Users with football knowledge showed a minimum increment in the understanding of the playing space. The geovisualization got 63 correct responses compared to 62 for the raw animation. This minimum increase is evident in the incorrect answers with the same number of responses for both animations (6). Participants aimed to answer the geovisualization animation more than the other, showing just one response compared to the two non-answered for the raw animation.

– *Participants without football knowledge*

Users without football knowledge had the same trend as the whole sample and the participants with football knowledge, with an increase in the correct answers from the raw animation to the Convex Hull animation. The Convex Hull animation got 35 correct answers compared to the 33 of the raw animation. The incorrect answers for the raw animation represent 10% of the sample, and 5% of the participants decided not to answer the question.

The Convex Hull animation shows fewer participants with incorrect answers (3%) but a slight increase in the non-answered questions (8%).

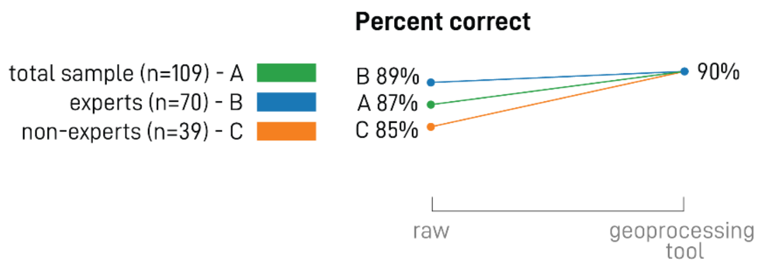


*Figure 15 Percentage of correct responses for the Convex Hull animation*

At a team-level analysis, the team's coverage space on the pitch can be analyzed through a Convex Hull. When a team is attacking, it tends to cover a larger pitch area. The identification of the coverage of a team when attacking seems to be easy to identify without the geoprocessing tool. There is a slight increment in the correct answers in both groups. Additionally, for the participants without football knowledge, the participants preferred not to answer the question when looking at the Convex Hull animation. Therefore, the visualization of the coverage area can be understood with a raw animation. Nevertheless, an in-depth analysis of the area a team covers can use the Convex Hull tool for detailed analysis.

## 4.3. Parameters of football analysis

There is a potential cartographic interest in analyzing football data based on its spatiotemporal nature. Football researchers are currently using geoprocessing tools to analyze football data in detail. For instance, the geoprocessing tools that are considered for the evaluation in this research have been used to make an in-depth analysis of different football parameters.

At a team-level analysis, researchers use parameters such as the *playing formation* of a team (Caldeira et al., 2022; Low et al., 2021; Shaw & Glickman, 2019), the *dominant region* of a team that is either attacking or defending (Fonseca et al., 2013), and the *playing space* that a team covers in different situations during the match (Moura et al., 2012).

For an analysis at a player level, researchers consider parameters such as the dyadic system of the *attacker-defender distance* (Bauer et al., 2022; Vilar et al., 2012), the *relative distance for a defender to intercept a shot* (Vilar et al., 2012), and the *distance between teammates* (Fonseca et al., 2013).

Nevertheless, the three tools evaluated in this research have been used for data analysis rather than visualization. This difference in the approach of the tools makes it necessary to emphasize the results obtained from this evaluation. As mentioned in the methodology, each parameter is the basis for the design of the evaluation questionnaire. Based on this logic, the results show a difference in the correctness of the answers for each parameter, which can be used for further research in cartographic football visualization.

### *4.4. Geoprocessing tools effectiveness*

This research evaluated the effectiveness of three geoprocessing tools: Nearest Distance calculation, Voronoi diagram, and Convex Hull. As a starting stage, four animations were designed to visualize 30 seconds of a real football game using a tracking football dataset. A user experiment was designed to evaluate the four geovisualizations. The method for evaluation was an online survey that collected information from 109 participants. The evaluation questionnaire considered the six parameters of football data analysis described in the theoretical background chapter.

Table 12 shows the difference between the correct answers obtained for the animation with the geoprocessing tool and the raw animation. Three parameters considered show a positive correctness in both groups, the *playing formation* with the Nearest Distance tool, the *attacker-defender distance* with the Nearest Distance tool, and the *playing space* with the Convex Hull. This positive correctness means that the geoprocessing tool used to promote the understanding, proved to be effective in these three parameters. The other parameters show a variation either negative or positive in the correctness of answers for the two considered groups, participants with football and without football knowledge. For the playing formation parameter, the Voronoi animation shows a decrease in 3 points in the correct answers for the participants without football knowledge. The Convex Hull animation shows that there is a decrease of 6 points in the correct answers for the users with football knowledge.

The Nearest Distance geoprocessing tool is ineffective for understanding the relative distance to intercept a shot that a defender must maintain. However, positive effectiveness is shown when visualizing dyadic systems at a player level, such as attacker-defender distance. The dominant region visualized by Voronoi cells had a positive effectiveness in the total sample and the participants with football knowledge, although a negative effectiveness for the users without football knowledge. To visualize the distance between teammates, the Voronoi animation shows a negative effectiveness. However, when analyzed by the knowledge group, the Voronoi animation had positive effectiveness for the participants with football knowledge and a negative effectiveness for the participants without football knowledge. The Convex Hull animation shows a positive effectiveness when visualizing the playing space of teams. Both knowledge groups had a positive increase in correct answers, particularly the participants without football knowledge.

<table>
<tr><th colspan="5">Table 12 Correctness for each geoprocessing tool by parameter analyzed</th></tr>
<tr><th rowspan="3">Geoprocessing tool</th><th rowspan="3">Parameter</th><th colspan="3">Correctness</th></tr>
<tr><th rowspan="2">total (n=109)</th><th colspan="2">football knowledge</th></tr>
<tr><th>with (n=70)</th><th>none (n=39)</th></tr>
<tr><td>Nearest Distance</td><td rowspan="3">Playing formation</td><td>3%</td><td>3%</td><td>3%</td></tr>
<tr><td>Voronoi</td><td>3%</td><td>6%</td><td>-3%</td></tr>
<tr><td>Convex Hull</td><td>0%</td><td>-6%</td><td>10%</td></tr>
<tr><td rowspan="3">Nearest Distance</td><td>Attacker-defender distance</td><td>7%</td><td>6%</td><td>10%</td></tr>
<tr><td rowspan="2">Relative distance to intercept a shot</td><td>-5%</td><td>-6%</td><td>-3%</td></tr>
<tr><td>-5%</td><td>-4%</td><td>-5%</td></tr>
<tr><td rowspan="2">Voronoi</td><td>Dominant region</td><td>1%</td><td>6%</td><td>-8%</td></tr>
<tr><td>Distance between teammates</td><td>-1%</td><td>1%</td><td>-5%</td></tr>
<tr><td>Convex Hull</td><td>Playing space</td><td>3%</td><td>1%</td><td>5%</td></tr>
</table>

The critical component of the evaluation was the design of the online survey. The survey design in sections helped to focus separately on the analysis and could potentially help further improvements. For instance, the demographic section was meant to help with an in-depth analysis of the results. Nevertheless, the results based on sex did not show a different trend than the analysis based on the participants' knowledge. Moreover, an analysis based on the age of the participants did not show significant differences since most participants ranged between the ages of 21 and 30.

The football knowledge section only considered the background of the user in an objective sense. The initial idea for identifying subgroups among the participants, based on their football knowledge, was not possible to use with the obtained results. Filtering the participants through each question did not show a significant number to separate the sample into subgroups. This subgroup analysis can be improved by including questions based on a subjective perception of the user's football knowledge. This consideration can help to add a much broader overview to understand the user perception of the animation. I used three questions regarding football knowledge to categorize the participants with football knowledge and without football knowledge. It is essential to consider adding more questions for a more detailed consideration of the user knowledge.

The online survey tool provided the flexibility necessary for the survey's design and the display of the animations. The possibility of sharing the survey online made it an easy tool to gather information from more participants. The consideration of a two-pilot test design proved effective for deploying the survey as all the participants completed the entire survey. The design of a specific feedback-oriented pilot test proved to be more effective for receiving feedback from the test participants. It facilitated the answers and guided the participants to give detailed comments on what they observed as a weak part that could be improved.

The evaluation section performed in a smooth connection of the pages that can be evident in the average time that the survey took for all the participants (12 minutes). Furthermore, the repeated questions for each animation created a sense of doubt for the correct answer to the user and might have created a learning effect during the survey. Future tests can include different game events rather than only one specific event.

The design of the evaluation section considered comparing the same questions for each type of geoprocessing tool with a raw animation. The repetition of the first question (*What is the lineup or formation of the red team?*) might have caused doubt in the participants if they answered correctly or not. Therefore, a learning effect might have occurred.

# 5. Conclusion

Research in football data focuses on using geoprocessing tools for data analysis rather than visualization. In this research, geoprocessing tools were focused on visualizing football tracking data and evaluating how effective these tools are for understanding specific football tactic parameters among users with and without knowledge of football. The geoprocessing tools analyzed were the Nearest Distance calculation, the Voronoi diagram, and the Convex Hull. Based on the literature review, I selected six parameters to evaluate users' understanding of football tactics. Researchers use the following parameters to analyze tactics at a team's level, such as the *playing formation* of a team (Caldeira et al., 2022; Low et al., 2021; Shaw & Glickman, 2019), the *dominant region* of a team that is either attacking or defending (Fonseca et al., 2013), and the *playing space* that a team covers in different situations during the match (Moura et al., 2012). To analyze at a player's level, researchers consider parameters such as the dyadic system of the *attacker-defender distance* (Bauer et al., 2022; Vilar et al., 2012), the *relative distance for a defender to intercept a shot* (Vilar et al., 2012), and the *distance between teammates* (Fonseca et al., 2013). These six parameters served to design the questions for evaluating the geoprocessing tools.

The online survey collected data from 109 participants. Participants with football knowledge represent 64% of the sample (n=70). The sample remaining 36% (n=39) represent participants without football knowledge. The evaluation analysis first considers the formation of a team because the question for the evaluation was the same for all the animations. It shows that the Nearest Distance calculation is more effective when users try to identify the playing formation of a team. Also, the Nearest Distance calculation proved to be more effective to visualize the attacker-defender distance. The Convex Hull animation proved effective for visualizing a team's playing space. The other parameters have results that vary from the groups analyzed; for instance, the Nearest Distance calculation showed to be ineffective for visualizing the relative distance to intercept a shot, having negative results for all the groups.

The Voronoi diagram shows variations, too. For visualizing the dominant region of a team, the total sample and the users with football knowledge show positive effectiveness, in contrast with the negative for the participants without football knowledge. At a player-level analysis, the visualization of the distance between teammates through a Voronoi diagram is negatively effective for the total sample, but has variations in both groups, the participants with football knowledge showed an increase by 1 point in effectiveness, and on the contrary, participants without football knowledge had a negative correctness.

The analysis and output of football tracking data and event data can be a matter of further consideration in a cartographic context. The animations generated for the evaluation are a scatter plot that displays the position of the players and the ball in each frame that the temporal resolution of the dataset allows, for the dataset used, 24 frames per second. Considering that a scatter plot is a tool that can be generated on different platforms, the visualization of football data can be of interest for further implementations and explorations on geovisualizations and user interactivity. Finally, the entire survey can be replicated to evaluate different events in a football match. With the same parameters considered, it can be a further detailed analysis to understand how effectively users visualize football data.

# Bibliography


Andrienko, G., Andrienko, N., Anzer, G., Bauer, P., Budziak, G., Fuchs, G., Hecker, D., Weber, H., & Wrobel, S. (2021). Constructing Spaces and Times for Tactical Analysis in Football. *IEEE Transactions on Visualization and Computer Graphics*, *27*(4), 2280–2297. https://doi.org/10.1109/TVCG.2019.2952129

Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., von Landesberger, T., & Weber, H. (2017). Visual analysis of pressure in football. *Data Mining and Knowledge Discovery*, *31*(6), 1793–1839. https://doi.org/10.1007/s10618-017-0513-2

Bauer, P., Anzer, G., & Smith, J. W. (2022). Individual role classification for players defending corners in football (soccer): Categorisation of the defensive role for each player in a corner kick using positional data. *Journal of Quantitative Analysis in Sports*, *18*(2), 147–160. https://doi.org/10.1515/jqas-2022-0003

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2014). Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. *2014 IEEE International Conference on Data Mining*, 725–730. https://doi.org/10.1109/ICDM.2014.133

Bradley, P., Ju, W., Laws, A., Gomez-Diaz, A., Martin-Garcia, A., & Evans, M. (2020, November). Football visualization: Capturing chaos and cultivating context. *Football Medicine & Performance*, 50.

Caldeira, N., Lopes, R. J., Fernandes, D., & Araujo, D. (2022). From Optical Tracking to Tactical Performance via Voronoi Diagrams: Team Formation and Players' Roles Constrain Interpersonal Linkages in High-Level Football. *Sensors*, *23*(1), 273. https://doi.org/10.3390/s23010273

Coito, N., Davids, K., Folgado, H., Bento, T., & Travassos, B. (2022). Capturing and Quantifying Tactical Behaviors in Small-Sided and Conditioned Games in Soccer: A Systematic Review. *Research Quarterly for Exercise and Sport*, *93*(1), 189–203. https://doi.org/10.1080/02701367.2020.1823307

Demaj, D. (2013). Geovisualizing spatio-temporal patterns in tennis: An alternative approach to post-match analysis. *Proceedings of the 26th International Cartographic Conference*. Proceedings of the 26th International Cartographic Conference, Dresden, Germany.

Fonseca, S., Milho, J., Travassos, B., Araújo, D., & Lopes, A. (2013). Measuring spatial interaction behavior in team sports using superimposed Voronoi diagrams. *International Journal of Performance Analysis in Sport*, *13*(1), 179–189. https://doi.org/10.1080/24748668.2013.11868640

Kotzbek, G., & Kainz, W. (2014). Football game analysis: A new application area for cartographers and GI-scientists? *5th International Conference on Cartography and GIS*. 5th International Conference on Cartography and GIS, Riviera, Bulgaria.

Liu, N. (2022). *Geovisualization of football players movement* [Diploma thesis, Erasmus Mundus Joint Master Degree Programme]. https://theses.cz/id/mbeof5/Thesis_Liu.pdf

Liu, N., Zhuang, L., Nétek, R., & Strobl, J. (2022). The overview, outcome, and outlook of GIS-based football visualization. *13th Geomoundus Conference*. 13th Geomoundus Conference, Lisbon, Portugal. https://geomundus.org/2021/docs/papers/Short_Paper_Geomundus_2021_Nianhua_Liu.pdf

Low, B., Rein, R., Raabe, D., Schwab, S., & Memmert, D. (2021). The porous high-press? An experimental approach investigating tactical behaviours from two pressing strategies in football. *Journal of Sports Sciences*, *39*(19), 2199–2210. https://doi.org/10.1080/02640414.2021.1925424

Metrica Sports. (n.d.). *Metrica Sports Sample Data*. Metrica Sports. https://github.com/metrica-sports/sample-data.git

Moura, F. A., Martins, L. E. B., Anido, R. D. O., De Barros, R. M. L., & Cunha, S. A. (2012). Quantitative analysis of Brazilian football players' organisation on the pitch. *Sports Biomechanics*, *11*(1), 85–96. https://doi.org/10.1080/14763141.2011.637123

Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A public data set of spatio-temporal match events in soccer competitions. *Scientific Data*, *6*(1), 236. https://doi.org/10.1038/s41597-019-0247-7

Perin, C., Vuillemot, R., & Fekete, J.-D. (2013). SoccerStories: A Kick-off for Visual Soccer Analysis. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2506–2515. https://doi.org/10.1109/TVCG.2013.192

Perin, C., Vuillemot, R., Stolper, C. D., Stasko, J. T., Wood, J., & Carpendale, S. (2018). State of the Art of Sports Data Visualization. *Computer Graphics Forum*, *37*(3), 663–686. https://doi.org/10.1111/cgf.13447

Shaw, L., & Glickman, M. E. (2019). *Dynamic analysis of team strategy in professional football*. https://api.semanticscholar.org/CorpusID:211526838

Stein, M., Janetzko, H., Lamprecht, A., Breitkreutz, T., Zimmermann, P., Goldlucke, B., Schreck, T., Andrienko, G., Grossniklaus, M., & Keim, D. A. (2018). Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 13–22. https://doi.org/10.1109/TVCG.2017.2745181

Vilar, L., Araújo, D., Davids, K., & Travassos, B. (2012). Constraints on competitive performance of attacker–defender dyads in team sports. *Journal of Sports Sciences*, *30*(5), 459–469. https://doi.org/10.1080/02640414.2011.627942

Wu, Y., Xie, X., Wang, J., Deng, D., Liang, H., Zhang, H., Cheng, S., & Chen, W. (2019). ForVizor: Visualizing Spatio-Temporal Team Formations in Soccer. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 65–75. https://doi.org/10.1109/TVCG.2018.2865041

# Appendices

Appendix A: Raw Animation
Appendix B: Nearest Distance calculation Animation
Appendix C: Voronoi diagram Animation
Appendix D: Convex Hull Animation
Appendix E: Online survey
Appendix F: Online survey results
Appendix G: Results key