

**Palacký University Olomouc, Faculty of Science,  
Department of Geoinformatics**

**Paris Lodron University Salzburg, Faculty of Digital and Analytical Sciences,  
Department of Geoinformatics**

# **GEOVISUALISATION OF FOOTBALL PLAYERS MOVEMENT**

**Diploma thesis**

Author

**Nianhua Liu**

Supervisor (Palacký University Olomouc)

Assis. Prof. dr. Rostislav NETEK

Co-supervisor (Paris Lodron University Salzburg)

Prof. dr. Josef Strobl

**Erasmus Mundus Joint Master Degree Programme  
Copernicus Master in Digital Earth  
Specialization Track Geovisualization & Geocommunication  
Olomouc, Czech Republic, 2022**



Palacký University  
Olomouc



With the support of the  
Erasmus+ Programme  
of the European Union

# **ANOTATION**

The aim of this thesis is to analyse, compile, and compare a set of football visualization and animation in different methods. The main part of the content establishes a systematic concept for all the football visualization charts, diagrams, statistic graphs, and animations. Spatial and attribute information of football datasets including positions, tracking, and tactical statistics is briefly introduced. The essential steps of visualizing football data through different technologies are described and important findings are shared. Then, the efficient workflow of football animations based on different technologies (e.g. Python, R, HTML5) is explained in detail, as well as evaluates the solutions of Tableau, ArcGIS, Adobe After Effects, and QGIS. Finally, the tactical, valuable, and understandable information to football coaches and players, delivered by Geovisualisation context, are interpreted and discussed.

## **KEYWORDS**

football, visualization, animation, tracking data, event data

Number of pages: 64

Number of appendixes: 1

*This declaration must be written on a separate sheet and hand-signed with the full name of the author stating the place and date of submission of the work (date and signature must be handwritten). If the author in work has followed a third-party work, this fact must be explicitly stated in the statement.*

This thesis has been composed by *Nianhua Liu* for the Erasmus Mundus Joint Master's Degree Program in Copernicus Master in Digital Earth for the academic year 2020/2021 at the Department of Geoinformatics, Faculty of Digital and Analytical Sciences, Paris Lodron University Salzburg, and Department of Geoinformatics, Faculty of Science, Palacký University Olomouc.

Hereby, I declare that this piece of work is entirely my own, the references cited have been acknowledged and the thesis has not been previously submitted to the fulfilment of the higher degree.

12.05.2022, Olomouc

Nianhua Liu

*I have countless words to express my appreciation. During my CDE studying, I have been to more than 20 countries in Europe, published 2 journal papers and 5 conference papers. Life could not be more interesting.*

*I am a super football fan. I have played football for more than 12 years. And I am a lifelong fan of Football Club Barcelona. Football should be artistic, like the Tiki-taka. When I watched this kind of art from Xavi, Iniesta, and Busquets a few years ago, I always think about one question – Could I draw this art?*

*Image football pitch as a map, and I could draw this art as I hope. That's my motivation for studying GeoVisualization.*

*Thank God, I met two perfect supervisors - Assis. Prof. dr. Rostislav NETEK and Prof. dr. Josef Strobl. They give me suggestions on what I should do, motivate me to keep working and help me collect the data I need. Thanks to Nelson Schäfer from intake 1, his master thesis impresses my work a lot. And thanks to my colleague in Salzburg and Olomouc. You are my stable power to move forward.*

# Palacký University Olomouc

Faculty of Science

Academic year: 2021/2022

## ASSIGNMENT OF DIPLOMA THESIS

(project, art work, art performance)

Name and surname: Nianhua LIU  
Personal number: R200695  
Study programme: N0532A330010 Geoinformatics and Cartography  
Work topic: Geovisualisation of football players movement  
Work topic in English: Geovisualisation of football players movement  
Assigning department: Department of Geoinformatics

### Theses guidelines

The main objective of the master thesis is to analyze, compile and compare a set of football visualization and animation in different methods. The student will establish a systematic concept for all the football visualization charts, diagram, statistic graph, and animations. He will consider football datasets including spatial and attribute information – positions, tracking, and tactical statistics. He also examines an efficient workflow for visualizing football data through different technologies (e.g. Python, R, HTML5), as well as evaluates the limits of football animation in Adobe After Effects. Finally, student will interpret and discuss the context of Geovisualisation used for tactical, valuable, and understandable information to football coaches and players. The student will attach all the collected datasets and all the animations to the thesis in digital form. The student will create a website about the thesis following the rules available on the department's website and a poster about the diploma thesis in A2 format. The student will submit entire text (text, attachments, poster, outputs, input and output data) in digital form on a storage medium and the text of the thesis in two bound copies to the secretary of the department.

Extent of work report: max. 50 pages  
Extent of graphics content: as needed  
Form processing of diploma thesis: printed  
Language of elaboration: English

#### Recommended resources:

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

Andrienko, N., Andrienko, G., Barrett, L., Dostie, M., & Henzi, P. (2013). Space Transformation for Understanding Group Movement. *IEEE Transactions on Visualization and Computer Graphics*, 19, 2169-2178. doi:10.1109/TVCG.2013.193

Beernaerts, J., De Baets, B., Lenoir, M., & Van de Weghe, N. (2020). Spatial movement pattern recognition in soccer based on relative player movements. *PloS one*, 15(1), e0227746. doi:10.1371/journal.pone.0227746

Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F. J., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *Int J Sports Med*, 28(3), 222-227. doi:10.1055/s-2006-924294

Kotzbek, G., & Kainz, W. (2014). Football Game Analysis: A New Application Area For Cartographers And GI-Scientists? Paper presented at the 5th International Conference on Cartography and GIS.

Kotzbek, G., & Kainz, W. (2015). GIS-Based Football Game Analysis – A Brief Introduction to the Applied Data Base and a Guideline on How to Utilise It. Paper presented at the 27th International Cartographic Conference, Rio de Janeiro / Brazil.

Larsson, P., & Larsén, K. (2001). The use of dGPS and simultaneous metabolic measurements during orienteering. *Medicine and science in sports and exercise*, 33, 1919-1924. doi:10.1097/00005768-200111000-00018

Supervisors of diploma thesis: **RNDr. Rostislav Nétek, Ph.D.**  
Department of Geoinformatics

Date of assignment of diploma thesis: **December 9, 2021**

Submission deadline of diploma thesis: **May 5, 2022**

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA GEOINFORMATIKY  
17. listopadu 50, 771 46 Olomouc  
-1-

L.S.

---

doc. RNDr. Martin Kubala, Ph.D.  
Dean

---

prof. RNDr. Vít Voženílek, CSc.  
Head of Department

# CONTENT

<b>LIST OF ABBREVIATIONS .....</b>	<b>7</b>
<b>INTRODUCTION .....</b>	<b>8</b>
<b>1 OBJECTIVES.....</b>	<b>9</b>
<b>2 METHODOLOGY.....</b>	<b>10</b>
2.1 Used Data .....	10
2.1.1 Open dataset .....	11
2.1.2 Tracking data .....	12
2.1.3 Event data.....	13
2.2 Processing Procedure.....	13
2.3 Used Software and Methods .....	15
2.3.1 Python.....	15
2.3.2 Web Technology.....	16
2.3.3 R .....	17
<b>3 STATE OF ART.....</b>	<b>17</b>
3.1 History of sports visualization in football .....	17
3.2 GIS-based football visualization and analysis .....	18
3.3 Football visual analysis with animations and videos .....	20
<b>4 STATIC FOOTBALL VISUALIZATION .....</b>	<b>22</b>
4.1 Statistic Information.....	22
4.1.1 Column chart: Compared teams or players .....	23
4.1.2 Bar chart: Longer labels for comparison .....	24
4.1.3 Line plot: Tracking rank change with Bump Chart.....	25
4.1.4 Radar chart: Specifically deal with player statistic .....	27
4.1.5 Others: Specific purpose.....	28
4.2 Spatial Information.....	29
4.2.1 Basic pitch map.....	29
4.2.2 Positioning map.....	30
4.2.3 Event map .....	34
4.3 Spatio-temporal Information.....	36
4.3.1 Grid and pixel map .....	36
4.3.2 Passing network map.....	39
<b>5 FOOTBALL ANIMATION CASE STUDIES .....</b>	<b>41</b>
5.1 Python.....	41
5.1.1 Creation: Visualized with Bokeh .....	42
5.1.2 Function: In-depth Analysis.....	48
5.2 Web Technology.....	49
5.2.1 Creation: HTML5, JavaScript, and CSS.....	49
5.2.2 Function: Interaction .....	52
5.3 R.....	53
5.3.1 Creation: Standardized with ggsoccer .....	53
5.3.2 Function: Data standardization .....	54

5.4 Others .....	55
<b>6 RESULTS .....</b>	<b>58</b>
6.1 Static Visualization.....	58
6.2 Animation .....	59
<b>7 DISCUSSION .....</b>	<b>62</b>
<b>8 CONCLUSION .....</b>	<b>64</b>
<b>REFERENCES AND INFORMATION SOURCES</b>	
<b>ATTACHMENTS</b>	



# LIST OF ABBREVIATIONS

<b>Abbreviation</b>	<b>Meaning</b>
AIX	Adobe Illustrator Exchange
API	Application Programming Interface
CRAN	Comprehensive R Archive Network
CSS	Cascading Style Sheets
FCB	Football Club Barcelona
FIFA	Federation Internationale de Football Association
GIF	Graphics Interchange Format
GIS	Geographic Information System
GPS	Global Positioning System
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
PC	Personal Computer
PNG	Portable Network Graphics
SVG	Scalable Vector Graphics
W3C	World Wide Web Consortium
XML	Extensible Markup Language

# INTRODUCTION

Playing in a specific space with limited time, football is considered as a sport dynamic process under spatio-temporal framework. Due to recent technological developments, large volumes of players' movement data can be collected from a football match. Join with the event and players' attribute data, a large pool of football match data is conducted. However, literature shows that the capacity to visualize and analyse the data in meaningful ways has not been matched by the ability to collect it. Increasing data volumes didn't play a full role. Therefore, the question becomes how to visualize and transform the data into useful information on coaches, players, and more stakeholders.

Static visualization is the most popular and well-developed method in football analysis. However, the taxonomy of static football visualization is not clear, even overlooked altogether. The first purpose of this thesis is to present and characterise the static football visualization thoroughly in a clear structure. By taking the statistical character of football games adequately into account, the first part of the content establishes a systematic concept for all the football visualization charts, diagrams, statistic graphs, and plots. Spatial and attribute information of football datasets including positions, tracking, and tactical statistics is briefly introduced, as well as some recommended open datasets. In this context, tracking and event data are described from a geospatial perspective. It is decidedly intended to sort out the applications of statistic, spatial, and spatio-temporal information. The contribution of Geographic Information System (GIS) technology and visualization method is highlighted.

Furthermore, static visualization cannot fulfill the dynamic visualized demand of current football analysis. Football animation is developing in recent years for better visualization. Players' locations, match events, and time is combined as the three most important parameters of football analysis. Especially the dynamic event that involves a time component can be visualized best with animations, as progressing time can actually be depicted with time, and doesn't require alternative methods of presentation like in static maps (Schäfer, 2021). Thus, coupling these parameters in football animations together with the toolset of different software can result in an advanced and intuitive medium, that allows football coaches and players to easily perceive and understand the depicted phenomena.

Previous studies still lack an introduction and evaluation of existing technologies in this early-stage subject. The primary objective of this charter is to provide an overview of creating the workflow, functions, and target users of three popular technologies are explained in detail, as well as evaluating some other potential solutions.

Finally, the tactical, valuable, and understandable information to football coaches and players, delivered by Geovisualisation context, are interpreted and discussed. This thesis contributes to the development of visualization and animation in football analysis, which is of great meaning in the case of Geovisualization.

# 1 OBJECTIVES

Football visualization and animation is not a popular or widespread topic in the geospatial community. A few individual experiments have examined the range of possibilities to visualize the football movement, but they are not organized systematically. Meanwhile, previous GIS research on football focuses more on analysis than visualization. Therefore, this thesis is aimed to establish a systematic football visualization concept in theoretical, as well as analyse, compile, and compare a set of football visualization and animation in different methods in practical. The range of possibilities to use GIS methods in football visualization and animation is examined. A multitude of datasets describing football player's statistic and their movement data in several matches are introduced as the basis to create the visualization and animation.

In order to explore the possibilities of static football visualization, three groups of visualization works are created and the applied workflow, encountered issues, and the user scenario, are documented. First, visualization work of statistic information is introduced to explore some popular football visualization works without the pitch map. A multitude of visualized examples for different applications is presented including charts, plots, diagrams, and statistic graphs. Second, some examples of spatial information are created to create to analyse how to highlight the players' location in visualization work. Additionally, supplementary timestamp datasets are added to visualize the players' performance over a certain period, under the spatio-temporal information subchapter.

With the increasing volumes of football data, the limitation of static football visualization comes with more challenges in visualizing dynamic football data. Thus, exploring different efficient workflows to create football animation is a critical objective of this thesis work. The essential steps to animate football data through different technologies are described and important findings are shared. The solutions built by Python, Web technologies, and R are introduced in detail. Each animation constitutes a unique target group and user scenario. Some other solutions built by Tableau, ArcGIS, Adobe After Effects, and QGIS are also evaluated. The limitations concerning these methods are also documented. Considering the data format, complexity of workflow, and required work time. With this, the capabilities, as well as limits of animating large number of football matches, can be shown.

Overall, the results of the work will allow football stakeholders to learn more about the possibilities and applications of visualization and animation. The theoretical output of this thesis draws a framework of football visualization concept to what methodology can be used for different purposes. In addition, the practically created set of animations can demonstrate the value to different users and can give the viewer an idea of what information can be delivered with the utilized tools and the established workflow.

## 2 METHODOLOGY

Utilized datasets, technologies, and processing procedures that were used to achieve the set objectives will be briefly introduced in this chapter. The first section provides a list of used data, including the open-source football datasets, as well as introduces their types. This is followed by a brief description of the conducted workflow with the most important aspects. Finally, the used software and tools are outlined. In the following chapters, these methods and procedures will be analysed more extensively.

### 2.1 Used Data

Football is influenced by two groups of factors — External factors and Match factors. External factors could affect the performance of football players. In this context not only personal factors such as mental, physical or medical conditions but also environmental factors such as weather, temperature, humidity, altitude, pitch conditions, stadium atmosphere, home and away game, or turf quality. Although atmospheric conditions undeniably have a certain impact on the players' performance, the measurement of the external factors' impact on the players is almost impossible Kotzбек and Kainz (2015b).

Some of the environment data can be obtained by Copernicus Services. For example, Copernicus Atmosphere Monitoring Service (CAMS) allows users to assess the past and predict the weather and air condition in the next few days. The temperature, sunshine, and humidity condition can be considered to analyse the players' performance. The Copernicus Land Monitoring Service (CLMS) provides geographical information on land cover and its changes. Long term turf quality of a football stadium can be monitored with CLMS. These services can provide tactical information to prepare for a football game.

However, the main content of this thesis deals with the visualization of match data. This kind of data should be captured and recorded for a specific purpose. Copernicus Earth Observation Programme doesn't provide such service related to football matches currently. Therefore, Copernicus data is not used in this thesis. External factors are not taken into account due to the lack of appropriate data.

Environment data is the objective condition of the football game, but it is not the decisive factor. The players' performance, including their shooting, passing, foul events, and movements are recorded as Match data. The football strategies are influenced by events variables and teams alter their playing style accordingly during the match. Match data can be quantified and evaluated to analyse individual and team performance (Peñas & A, 2010). Understanding the players' performance is the key to winning a football game. This thesis focuses on the visualization of match data with geospatial information.

Several types of data were identified to use in sports visualizations: box score data like statistical summaries, tracking data about in-game actions and trajectories, and meta-data about the sport and its participants (Perin et al., 2018). Event data and tracking data are two types of geo data regarding football games. Research shows that more practitioners use event data than tracking data (Herold, Kempe, Bauer, & Meyer, 2021). Both can be regarded as vector data in form of point features with attribute information. Abstracting from the various specific technologies for collecting movement data, they can be recorded by time-based, change-based, location-based, and event-based methods (N. Andrienko, Andrienko, Pelekis, & Spaccapietra, 2008).

### 2.1.1 Open dataset

Football-specific geo data was referred to a special kind of geo data that encompass all spatio-temporal data which directly represent the entire gameplay of a football game (Kotzbek & Kainz, 2015b). A typical data set for one football game consist of general match information, team information about the players on the pitch and substitutions, positional data, and event-specific characteristics (G. Andrienko et al., 2021).

Football datasets are usually made available in the form of, for example, CSV tables. For an individual match, event datasets have several hundred kilobytes in size, while tracking data can reach dozens of megabytes on average. Many open football datasets are available online. However, most providers like Metrica Sports, Wyscout, or StatsBomb charge money to get access to their datasets. But they also provide some trail data or open datasets through Github and other platforms. Some recommended open datasets are listed:

#### **StatsBomb Open Data**

Available at <https://github.com/statsbomb/open-data>.

StatsBomb provides the largest open football data online. The data from more than thousands of matches are provided as JavaScript Object Notation (JSON) files exported from the StatsBomb Data Application Programming Interface (API), including data like Competition and seasons, Matches, Events, and lineups for each match, respectively. Each file is named for a match ID. Some documentation about the meaning of different events and the format of the JSON can be found in the doc directory. A wide-used football analysis Python library mplsoccer has integrated the Statsbomb open data (StatsBomb, 2022).

#### **Kaggle**

Available at <https://www.kaggle.com/datasets/secareanualin/football-events>.

Kaggle Football Events provides the main dataset with 11 types of events, a granular view of 9,074 games, totalling 941,009 events from the biggest five European football leagues from 2011/2012 season to 2016/2017 season (Kaggle, 2017). Over 90% of the played games during these seasons have event data. This dataset is a result of web scraping and integrating different data sources. The central element is the text commentary. All the events were derived by reverse engineering the text commentary, using regex. The dataset is organized into three files:

events.csv - contains event data about each game. Text commentary was scraped from: bbc.com, espn.com, and onefootball.com

ginf.csv - contains metadata and market odds about each game. odds were collected from oddsportal.com

dictionary.txt - contains a dictionary with the textual description of each categorical variable coded with integers

#### **Wyscout: Spatio-temporal match events**

Available at <https://doi.org/10.6084/m9.figshare.9711164>.

Available at <https://doi.org/10.6084/m9.figshare.c.4415000.v5>.

This public scientific research dataset describes the largest open collection of football logs ever released, collected by Wyscout, containing all the spatio-temporal events (passes, shots, fouls, etc.) that occurred during each match for an entire season of seven prominent football competitions. Each match event contains information about its position, time, outcome, player, and characteristics (Pappalardo et al., 2019).

The football-logs have been collected and provided by Wyscout. The procedure of data collection is performed by video analysts, who are trained and focused on data collection for football, through a proprietary software. The tagger has been developed and improved over several years and it is constantly updated to always guarantee better and better performance at the highest standards.

### **Metrica Sports Sample Data**

Available at <https://github.com/metrica-sports/sample-data>.

Metrica Sports provides open sample data through Github, including sample tracking and event data. Tracking and event data are synchronized. Two games are provided in standard CSV format, and one game is provided in JSON, Extensible Markup Language (XML), and txt file. The data of these games is anonymized, meaning there are no references to the names of players, teams or competitions. The data goes from 0 to 1 on each axis. The coordinate (0,0) is the top left, (1,1) is the bottom right, and (0.5,0.5) is the kick off point. The dimensions of the field are the same for both games: 105x68 meters (Metrica Sports, 2022).

### **FBref: Football Statistics and History**

Available at <https://fbref.com/en>.

FBref has the largest Easy-to-use source for football statistics data including player, team, and league. The football data is provided by Data Sports Group, while the advanced statistics is provided by StatsBomb. FBref is exhaustive in statistics domain. It's convenient to collect the statistics data through this website.

## **2.1.2 Tracking data**

Tracking data can be described as continuous paths in form of Moving Point Objects (MPO), which is widely used within several different GIS applications (Dodge, Weibel, & Lautenschütz, 2008). These are characterised as point clouds, and its elements are related to each other and are chronologically ordered (Frentzos, Theodoridis, & N. Papadopoulos, 2009; Sester, Feuerhake, Kuntsch, & Zourlidou, 2015).

Tracking data in football corresponds to the capture of a vast amount of georeferenced data digitalized the players' and the ball's movement as consecutive point data (Carling, Williams, & Reilly, 2006). Specifically, football matches comprise rich, multivariate movement data at high temporal and geospatial resolution (Manuel Stein et al., 2015). Tracking data mirror the players' and the ball's trajectories, which correspond to point features in the form of time dependent X-Y coordinates. The provision of the z-coordinate is also common for the ball's path in order to reconstruct 3D football trajectory (Memmert, 2019). Speed can also be calculated at each single point and also provided as an attribute. All this information is separately available for each player as well as the ball. (Kotzbek & Kainz, 2015b). As similar patterns may occur at different spatial locations, constructing new spatial reference systems for aligning and matching movements irrespective of their absolute locations is necessary for football tracking data (G. Andrienko et al., 2021). Instead of "tracking data", the term "positioning data" is also common.

Tracking data are usually extracted from Global Positioning System (GPS) devices or video recordings. Wearables in electronic performance and tracking systems are approved from Federation Internationale de Football Association (FIFA) in 2015 (Pettersen, Johansen, Baptista, Halvorsen, & Johansen, 2018). Lots of stadiums are equipped with stationary installations of multiple cameras that record games from different viewpoints. Video analysis software is used for extracting time-stamped positions of the players,

referees, and the ball from video footage, usually with a sampling rate of 10-25Hz (Csanalosi, Dobreff, Pasic, Molnar, & Toka, 2020).

Tracking data can be correlated with the event data as an attribute. Comparable to event data, tracking data in football corresponds to consecutive point data linked with a dense timestamp. Whereas tracking data is directly linked to the players and the ball, event data represents connected dynamic interactions among the players.

### **2.1.3 Event data**

The players' motion has to be distinguished from their movement. Whilst the former is associated with the objects' spatial variance, the latter corresponds to its actions (Kang, Hwang, & Li, 2006). The position-based specific behaviours of football players on the field are called events. Event data are a sequence of significant events that occur during a match. Events can be broadly categorised as ball events and technical events. Ball events contain player relevant information such as passes, tackles, corners, free kicks, throw-ins, goalie kicks, penalties, and shots. Technical events include fouls, time-outs, offside positions, substitutions, or cards. Each row of event data should contain information about the time, space, specific players, and event characteristics. Event data are not dense as tracking data qualitatively. Samples are only captured when an event occurs. However, they can be semantically richer as they include details like the type of event and the players involved (Gudmundsson & Horton, 2017).

Event data are usually extracted from video recordings. Positions and times of the events are recorded as position based attributive data. Unfortunately, even the newest video algorithm cannot understand the specific event correctly. That's why event data is captured partly manually from video by so called scouts, who gather the data while repeatedly watching every game scene of the recorded video. 1,500 to 3,000 event data records per one single game are common. These data are also classified, with more than 60 categories including manifold forms available. Therefore, an event data set for one single game is highly informative and provides a valuable dataset from which detailed analysis can be performed (Kotzbek & Kainz, 2014).

Due to the fact that the data capture procedure depends on individuals' subjective perceptual processes, these data are susceptible to errors. To minimize the expected error, all records are cross validated by one other person. Some machine learning approaches that facilitate manual annotation are also used. An average error of 3.6 m was observed but the reliability is still sufficiently reliable for the coaching context (Bradley, O'Donoghue, Wooster, & Tordoff, 2007). The key to overcoming the comparatively high spatial error problem is a joint connection between tracking and event data. Since it is common practice that each point is equipped with a consecutive time tag and frame, tracking and event data can be attributively joined (Kotzbek & Kainz, 2018).

## **2.2 Processing Procedure**

Before the practical implementation work, a considerable amount of literature is read and concluded to consolidate related works on the football visualization and animation. It takes time to learn and understand types, approaches, and visual styles on this topic. Based on this knowledge, the set of available tools and technical solutions were explored to create static visualizations and animations work.

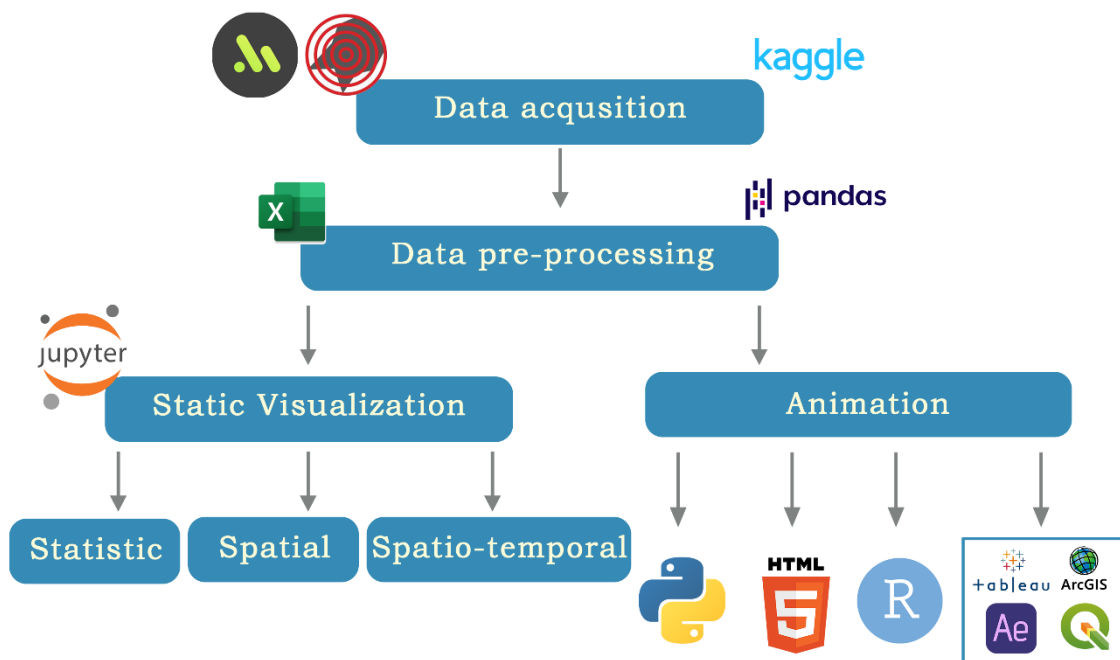
The acquisition of data is conducted online via Github and football data portals. Relevant open datasets were collected and pre-processed in Excel and Python Pandas library. The straightforward processing process must include all necessary steps. Here,

the processing steps include calculations of statistics, filtering of the data, and spatial analysis. Statistic visualization work without a pitch map is created based on the pre-processed data. Multiple Python libraries in Jupyter Notebook are explored to create charts, plots, and diagrams for different application purposes. Functionalities that looked promising to be implemented in the process are documented and practiced.

Before creating some spatial and spatio-temporal visualization, coordinates transformation and drawing a pitch map are necessary steps. The coordinates transformation is based on determining the relative position points of all trajectories with respect to the corresponding point of a chosen or constructed reference trajectory and its movement vector (Natalia Andrienko, Andrienko, Barrett, Dostie, & Henzi, 2013). The corresponding points will be projected to a football pitch. The total pitch size of the international standardised football field is determined to 105m \* 68m by the FIFA laws. However, the origin of the local coordinate system is not standardized. Some of the local coordinate systems positioned the origin at the field's center, while the other put it at one of the corner points. Thus, all the data acquired should be pre-processed based on the corresponding pitch coordinate system.

Next, the spatial and spatio-temporal visualizations based on the pitch are created. They are distinguished due to the difference in highlight points. Spatial visualization work emphasizes players' location, while the specific location can be observed and described. Spatio-temporal tends to evaluate the performance of a player at a certain period, while the specific location of an individual player cannot be described at a certain moment. Analysis of tracking and event data can reflect specific functions of pitch regions (Gudmundsson & Horton, 2017). Related operations, applied workflow, encountered issues, and user scenarios, are documented.

Then, with different programming tools, graphic software, and installed plug-ins, animations are conducted. The finished animation compositions were then exported as Graphics Interchange Format (GIF) for presentation. In the following subchapters 5.1 to 5.4, the workflow, function, and target users are explained in detail by the examples of the created football animations. The general processing workflow is shown in Fig. 1.



**Figure 1** Workflow of thesis



## 2.3 Used Software and Methods

### 2.3.1 Python

Python is the most frequently used programming language in this thesis. The results of the static football visualization part are all created in Python environment. As a high-level and interpreted language, Python consistently ranks as one of the most popular programming languages. With a large developer community, Python has a pool of comprehensive standard libraries.

#### **Jupyter Notebook**

Jupyter Notebook, based on Python 3, is the software used for Python programming in this thesis. Jupyter Notebook is the latest web-based interactive development environment, with the open-source application that allows researchers to create, share and visualize code. It runs in the browser and provides a simple lightweight yet powerful platform for scripting. For the current study, Jupyter Notebook was used to create and test different parts of the code as ipynb.

#### **Pip**

Pip is an easy tool to install the package of python packages. Programmers use pip to install packages from the Python Package Index and other indexes. All the python packages mentioned in the technologies were installed through pip in Jupyter Notebook.

#### **Pandas**

Pandas is a fast, powerful, flexible, and easy-to-use package that allows for efficient data manipulation and analysis, built on top of the Python programming language.

#### **NumPy**

NumPy is a python library that offers comprehensive mathematical functions and high-level functionality. For this thesis, NumPy was employed to execute mathematical formulas in the code. For example, feature vstack is used to create the arrays, and feature where is used to return elements depending on condition. Lots of other mathematical features are used here such as sqrt, round, and abs.

#### **SciPy**

SciPy is a scientific computation library that extends NumPy by providing additional tools for array computing and provides specialized data structures, such as sparse matrices and k-dimensional trees. The Convex hull in this thesis is created by the spatial module inside the SciPy library.

#### **Pylab**

Pylab is a convenient module that imports functions from the modules Numpy and Matplotlib in a single namespace. Numpy provides efficient numerical vector calculations based on underlying Fortran and C binary libraries. Matplotlib contains functions to create visualizations of data.

#### **Matplotlib**

Matplotlib is a plotting library for creating static, animated, and interactive visualizations in Python programming language and its numerical mathematics extension NumPy. In this thesis, Matplotlib can be used for drawing a pitch.

#### **Seaborn**

Seaborn is a Python data visualization library that uses Matplotlib underneath to plot graphs. It provides a high-level interface for drawing attractive and informative statistical graphics. It will be used to visualize random distributions.

### **Bokeh**

Bokeh is a Python library for creating interactive visualizations for modern web browsers. Beautiful graphics and a smooth interface for animation are built with streaming datasets. With Bokeh, users can create powered visualizations powered by JavaScript without writing any JavaScript script.

### **Functools**

Functools library is imported for the partial functions. Partial functions allow fixing certain values set of a function and generate a partial function application. Some portion or keywords of a function's arguments are frozen resulting in a new object with a simplified signature.

### **mplsoccer**

mplsoccer is a Python library for plotting football charts in Matplotlib and loading StatsBomb open-data.

## **2.3.2 Web Technology**

In this thesis, one of the animation solutions is built by web technology, including HyperText Markup Language 5 (HTML5), Cascading Style Sheets (CSS), and JavaScript. The whole web technology solution will be delivered as web interfaces structured in HTML5, styled with CSS, and its behaviour controlled by JavaScript.

### **HTML**

HTML is a core technology for building websites, which provides the structure of a web page and its content, by designing the function of each element in a page. HTML consists of a series of elements, which use to enclose, or wrap, different parts of the content to make it appear a certain way, or act a certain way. HTML allows web users to create and structure sections, paragraphs, and links using elements, tags, and attributes. However, it's worth noting that HTML is not considered a programming language, but an official web standard, as it can't create dynamic functionality. The World Wide Web Consortium (W3C) maintains and develops HTML specifications, along with providing regular updates.

### **CSS**

CSS is a style sheet language used for describing the presentation of a document written in a markup language such as HTML, which is designed to separate the presentation and content. CSS describes the aesthetics of the layout, defining colour, sizes, font, and element positions.

### **JavaScript**

JavaScript is a scripting programming language for implementing functions into websites (Wirfs-Brock & Eich, 2020). JavaScript is used by most pages that contain dynamically updated content and interactive features, including interactive maps, animated 2D and 3D graphics, video players, and similar features. With its functionality, it complements the core web technologies: HTML and CSS.

JavaScript contains lots of useful libraries such as jQuery. jQuery is a fast, small, and feature-rich JavaScript library with a combination of versatility and extensibility. It is designed to simplify HTML DOM tree traversal and manipulation, event handling, animation, and Ajax. It makes things much simpler with an easy-to-use API that works across a multitude of browsers, which results in the most widely deployed JavaScript library (jQuery, 2022).

### 2.3.3 R

R is an open-source interpreted programming language for a wide variety of statistical computing and graphics techniques. Users have created thousands of packages to augment the capabilities of the R language (R Foundation, 2022). RStudio provides open-source programming software for R, which integrates the development environment (RStudio, 2022). In this thesis, R is used for creating one of the animated solutions.

#### **ggplot2**

ggplot2 is an R package dedicated to data visualization, which is a system for declaratively creating graphics. ggplot2 is a powerful set of tools for visualising and investigating data. It can greatly improve the quality and aesthetics of the graphics (ggplot2, 2022).

#### **ggsoccer**

ggsoccer is an R package specialising in football visualization. ggsoccer is implemented as a layer of an open-source data visualization package called ggplot. It provides a set of functions, extensible layers, and themes that make it easy to plot and customize soccer event data in R environments. It is designed to work smoothly with a variety of popular sports data providers (ggsoccer, 2022).

#### **gganimate**

gganimate provides a range of new functionality that can be added to the plotted object and customize how it should change with time. gganimate extends the grammar of graphics as implemented by ggplot2 to include the description of animation (gganimate, 2022).

## 3 STATE OF ART

### 3.1 History of sports visualization in football

Sports have a long history of data collection and visualization. In the 1870s, Henry Chadwick collected statistical information about baseball games (Dickson, 1997). Systematic analysis of sports play has been occurring since the 1950s using manual notation methods (Reep & Benjamin, 1968). For football analysis, the dependent variables more frequently used were football performance models from all comparative studies (Sarmiento, Marcelino, Campanico, Matos, & Leitão, 2014). Only a few studies have focused on football visualization.

Unfortunately, conventional methods of post-game analyses often fail to analyse and visualise the complex spatio-temporal patterns of the football sufficiently (Kotzbek & Kainz, 2014). Sports data tends to be hypervariate, temporal, relational, hierarchical, or a combination thereof, which leads to some fascinating visualization challenges (Basole & Saupe, 2016). Sports visualization offers new approaches to exploring, making sense of, and communicating sports data. The basic purpose of sports visualization is to convey knowledge and to display understandable and analysable information with tactical value. As visualizations can be more accessible and more meaningful than traditional statistical analysis (Janetzko et al., 2014).

The top leagues and clubs are aware of the necessity of football visualization and analysis, even though the potential of using them is not yet tapped. After FIFA's 2015 approval of wearables in electronic performance and tracking systems, the performance and tracking data from teams and players are increasingly being collected during practices and matches (Pettersen et al., 2018). However, sports data cannot release its

value without good visualization methods. Making the complex spatio-temporal tracking and event data understandable is so usable for professionals (G. Andrienko et al., 2021).

Except for the players, coaches, referees, sports analysts, and other professionals, the objects of competitive sports visualization can also include audiences and sponsors. Different objects have different desires for competitive sports data. In fact, print and online media companies emphasize football visualization more than clubs over the past decades. To attract more readers, television broadcast and betting websites increasingly use statistics and infographics to present sports data and predict sports events.

Existing football visualization research stems from passionate and distributed communities, blogs, and websites, working on sports data analysis and communication. With the increasing interest in football visualization from both practitioners and professional clubs, there is a need to document and analyse these efforts. While academia has completed lots of studies about sports visualizations, football as an application domain has not yet been the subject of a review from a Geovisualization perspective.

### **3.2 GIS-based football visualization and analysis**

Although the acquisition of football-specific geo data started from mid-1990s, the publications are still rare in the domain of cartography and GIS (Kotzbek & Kainz, 2015a). However, a series of research which is about GIS-based football game analysis and visualization are increasing. From a geographical perspective, football can be considered as a spatio-temporal framework. Spatio-temporal data is primarily used in team sports analysis (Gudmundsson & Horton, 2017). Within this scheme, spatio-temporal phenomena and objects are correlated (Kotzbek & Kainz, 2014). Specifically, this correlation is described in two categories: Visualization and Analysis.

Systematic overviews of the visualization methods of geospatial data analysis in general and for movement data in particular are presented before (N. V. Andrienko & Andrienko, 2005; Gennady Andrienko et al., 2013). A visualized solution for football player analysis was used in information visualization techniques, applications apply such as glyphs, modified star plots, details on demand, color, and gestalt principles. The analysis of large sums of football statistics can be extremely difficult and confused without graphical data and clear structure (Rusu et al., 2010). To better understand this, three main tasks for visualization were proposed: feature presentation, feature comparison, and feature prediction (Du & Yuan, 2021). And the most important feature of football is the players and ball movement.

Based on the players' movements, assessments of their behaviour can be drawn (Sester et al., 2015). It has to be taken into account, that in general these are influenced by environmental factors as well as other objects' impacts (Dodge et al., 2008). In this context, the trajectories reveal not only the individual, but also the collective behaviour (Gudmundsson, Laube, & Wolle, 2012; Kang et al., 2006).

The analysis of 2D or 3D trajectory football movement data is a task within geo-temporal data analysis. After simplification and aggregation, the spatio-temporal trajectories of football players can be represented in detail (Sacha et al., 2017). To support the interpretation and understanding of movement patterns, a dynamic approach that combines trajectory simplification and clustering techniques was presented (Sacha et al., 2017). Depending on the trajectory, timestamps, and potentially multivariate attributes recorded along the trajectory, numerous football movement features can be derived pertaining to curvature, direction distribution, speed, acceleration, etc. The complexity of

football movement data has been demonstrated with 11 categories, provided by several sports data collection companies (C. Perin, R. Vuillemot, & J.-D. Fekete, 2013).

All football data contains spatio-temporal information and statistical information. Lots of related tactical analysis were conducted based on two football taxonomy. The first one is based on the event of football, such as shots, passes, crosses, and tackles. The second one is based on the visualization perspective, such as graphs, charts, and plots (C. Perin, R. Vuillemot, & J. Fekete, 2013).

Visualization of team sports has recently been surveyed by a fully spatio-temporal lens (Gudmundsson & Horton, 2017). Large scale team analysis over a full season of player tracking data was enabled for both individual player and team analysis, providing context to player statistics (Bialkowski et al., 2014). Spatio-temporal tools such as passing speed calculation, Motion model and passable area, Pass sequence analysis, Clustering movement, and Correlating clusters were analysed in previous research (Gudmundsson & Wolle, 2014).

Pass analysis is the most basic one among all football events. As space is crucial in football, for each pass the analysis takes the game's current spatio-temporal structure into account in order to determine its inherent risk. The practicability of GIS addresses the problem of how to analyse single players' passes (Kotzbek & Kainz, 2017). To support comparisons between the maps of the risk pressure on the players, several pressure levels are represented by isolines. Each isoline connects points with the same level of pressure. (Gennady Andrienko et al., 2017). Two-dimensional kernel density plots showing Expected Contribution to the Outcome of the Match (ECOM) per 90 min and pass accuracies for goalkeepers, defenders, midfielders, and forwards (Bransen, Haaren, & Velden, 2019). Statistical properties of position-dependent ball-passing networks in real football games are examined, based on a Markov chain (Narizuka, Yamamoto, & Yamazaki, 2013).

A general framework with a feature model and several analysis methodologies were provided based on spatial and spatio-temporal analysis of football matches (H.-C. Kim, Kwon, & Li, 2011). Spatio-temporal patterns of football game were extracted to elaborate normative models of positioning behaviours measures (Gonçalves et al., 2019). Qualitative Trajectory Calculus was explored for spatial movement to recognize the pattern that occur on different parts of the field and/or at different spatial scales (Beernaerts, De Baets, Lenoir, & Van de Weghe, 2020). With the automated quantitative analysis of space dynamics, the capacity of evaluating space occupation and generation can be applied in specific matches and situations, and directly integrated into coaches' analysis (Fernández & Bornn, 2018). An alternative method to study the spatial relation established between players at each instant of a game is the Voronoi diagram, which is a spatial construction that allows a spatial partitioning of the field area into cells, each associated with each of the players, according to their position (Voronoi, 1908).

The most popular spatial analysis tool in previous football research is network analysis. The network analysis has already been applied to football games such as a structural property of ball-passing networks (Yamamoto, 2009; Yamamoto & Yokoyama, 2011), and assessment of players (Duch, Waitzman, & Amaral, 2010; Peña & Touchette, 2012). Football can be regarded from the network perspective as a competitive relationship between two cooperative networks with dynamic network topology and the dynamic network node. Many complex large-scale networks have been shown to have topological properties in common, based on a small-world network and scale-free network models (Yamamoto & Yokoyama, 2011). A weighted and directed network in which nodes correspond to players and arrows to pass. The resulting network or graph provides a

direct visual inspection of a team's strategy, from which play pattern, determine hot-spots on the play and localize potential weaknesses can be identified (Peña & Touchette, 2012).

The team is considered a network with players as nodes and passes as (directed) edges. To analyse the pass network behaviour at the complex network level, a temporal analysis of the resulting passes network is done, looking at the number of passes, length of the chain of passes, and the centrality of players on the turf (Cotta, Mora, Molina, & Merelo Guervós, 2011). Network work analysis has real tactical value. For example, different metrics of network analysis are combined to extract the particular signature of the F.C. Barcelona coached by Guardiola (Buldú, Busquets, Echevoyen, & Seirul.lo, 2019). The defenders of a football team for their spatial formations during a game were analysed (H.-C. Kim et al., 2011). In fact, the nature of these analysis tools and their usage suggests that these visualizations were not created by experts of geodata, but rather by professional football analysts who do not have specific expert knowledge on visualization made for GIS purposes

Only a few football analysis and visualization tools were also developed in the GIS domain (Liu, Zhuang, Néték, & Strobl, 2021). An approach for automated analysis of scoring attempt patterns utilising GIS was already presented (Kotzbek & Kainz, 2016). A GIS dynamic zoning approach of football pitches based upon the players' movements was developed based on Python and executed as a custom tool in ArcGIS for Desktop 10.x (Kotzbek & Kainz, 2018).

### **3.3 Football visual analysis with animations and videos**

Cartographers have been perfecting the representation of dynamic spatio-temporal phenomena with static, spatial representations in the form of two-dimensional maps for a long time (Harrower & Fabrikant, 2008). However, static representation is not a proper method for visual analysis of moving objects, such as football. Visual Analytics method presents a flexible, modular, and expandable layer-based system allowing in-depth analysis. Generally, an advantage of visual analysis is that these methods may be adapted more easily to changing user tasks (Manuel Stein et al., 2015). The integration of Visual Analytics techniques into the analysis process enables the analyst to find interesting events and communicate the found results. The feedback loop in the Visual Analytics pipeline helps to further improve the classification results (Janetzko et al., 2014).

Also, visual analysis of team sports data has recently gained interest, with a number of works addressing the football case. For example, a carefully crafted system for interactive visualization were introduced to present and summarize game situations in football matches (C. Perin et al., 2013). feature-based techniques were described to visualize, segment, and classify football data (Janetzko et al., 2014). Visual exploration of football player interactions and free space situations in an interactive system were analysed (Shao, Sacha, Neldner, Stein, & Schreck, 2016; M. Stein et al., 2016). A visual analysis approach for sketch-based visual search and exploration of football trajectories was presented to discover crucial game situations (Shao et al., 2016). An approach to defining and combining visual abstraction techniques interactively to overcome overplotting and clutter was presented (Sacha et al., 2017). A visual analysis system for interactive identification of football patterns and situations was proposed to automatically determine interesting game situations (Manuel Stein et al., 2015).

There are two types of visual analysis methods that are mainly used in football events — Video and Animation. Videos are mostly used in post-game analysis. Video analysts seek and identify key scenes, based on their experience. These key scenes are selected to

be representative and characteristic of a team's general behaviour. Normally, this analysis takes up to three working days. There are three major application areas of football video analysis: video summarization, provision of augmented information, and high-level analysis. Each of them delivers a different tactical value (D'Orazio & Leo, 2010). Football video segmentation systems were also described in previous research (Maćkowiak, Konieczny, Kurc, & Maćkowiak, 2010).

An experiment was performed to contrast different animations and video codings in terms of their cognitive and emotional effectiveness for spectating a football game. The conclusion is that more advanced renderings based on animation might be able to give an individually optimized blend (Wikstrand & Eriksson, 2002). Field, Tracking, and Video Editor tool provide an animation feature to help understand the coach intentions regarding a certain play or movement (Rodrigues, Cardoso, & Rodrigues, 2013). However, the research is rare in the domain of animation compared to the video.

In cartography, two basic animation types are known: temporal animation and non-temporal animation. Temporal animation deals with the depiction of dynamic events in chronological order, such as football games (Dransch, 2014). Animation was also viewed as the main screen of the event detecting prototype, which includes a football pitch. The pitch and the events are displayed synchronously linked to the animation (Gudmundsson & Wolle, 2010). Map animation was used as a tool for game episodes in static and dynamic modes for player-to-player pressure research at visual displays (Gennady Andrienko et al., 2017).

The knowledge from previous research was used to derive in this thesis developed workflow. The core concept of this master's thesis was to display movement data with Geovisualization and animation methods. There is a lack of clarity as to whether these methods are efficient solutions. Since no scientific publications exist, the selected appraisal cannot be evaluated by a peer-reviewed quantitative framework and is therefore to be assessed critically.

## **4 STATIC FOOTBALL VISUALIZATION**

Data visualisation is a tool that translates trends effectively using graphics (Tracey L. Weissgerber, Milic, Winham, & Garovic, 2015). The most potent mode to interpret findings is presenting as aesthetically pleasing forms of visuals, including graphical figures, charts, plots, and animation (Tracey L. Weissgerber et al., 2015; T. L. Weissgerber et al., 2019). These eye-catching visualizations look unique, slick, and beautiful to stimulate curiosity and attract people, which can selectively draw attention to relevant information.

However, incorporating aesthetics in visualization without losing accuracy and credibility can be hard. It's important to maintain a good balance between data mapping, interactivity and aesthetics (Bartels, De Buyser, & Van Ussel, 2013). This is especially evident in football, as coaches are expected to analyse each game using data and animation. Applying data visualisation to a football context is easier in theory than in practice.

In fact, there are thousands of graphs and charts for football visualization. This represents a risk of creating the problem of information overload, thus care must be taken to prevent already busy coaches from being exposed to an avalanche of data and video to view and interpret (Bradley et al., 2020). Therefore, static visualization has become popular as the amount of information has increased exponentially, while the time to interpret it has decreased (Lacome, Simpson, & Buchheit, 2018). Understanding the tactical value and workflow of static football visualization is necessary.

A thorough framework of static football visualization will be given in the following three subchapters. Each subchapter is dedicated to one of the three categories. They are applied in different user scenarios which provides different levels of information and tactics. First, in a short introduction, the content and useful background information is given for each of categories. In addition, the tactical value that can be derived from each static visualization is presented. Subsequently, the workflow from the very beginning of acquiring the data until the final export of the static visualization is elaborated. In the process, noteworthy tips as well as problems and their according solutions are presented in detail.

It should be noted that the chapters of the workflows of the static visualization build on one another. Many work steps and principles were used in the work repeatedly, like creating a pitch map. When they appear for the first time, they are explained in detail. But the second time they appear in the workflow, they are only briefly dealt with.

### **4.1 Statistic Information**

Statistic data presented in graphs, charts, plots are easy to grasp, but it is essential to find the right way to present the data (Azzam, Evergreen, Germuth, & Kistler, 2013). With a history of more than one hundred years, there are plenty of ways to represent data utilised in football. One way to do this is non-spatial statistic graphs and charts. Elite football games consist of 22 players, and this could limit what kind of charts are suitable. Displayed without a pitch map is the most notable feature of non-spatial statistic visualization. Even though this might damage the delivery of spatial information, it also provides a specific tactical value in post-game analysis.

Statistics visualizations are effective at better understanding a game. A live game might be too dynamic to discern and conclude. Statistics visualizations are often applied in post-game analysis instead of live-game analysis. They often contain complicated information which required time for interpretation.



For example, a coach never shows the football players a pie chart to explain the tactics at half-time or before substitution. Non-spatial graph is hard to convey any spatial information directly, which may damage its tactical value and lose some key information. It is difficult for players to understand the tactical instruction immediately.

A correct user scenario for statistics visualizations is retrospective analysis. When a football match is over, the coach might hope to compare the covered distance, completed passes, possession percentage, and other indexes with previous matches. These indices could be compared by statistics visualization directly. Besides, a game shown on television might be too dynamic for an audience to be able to discern everything that is happening in real time; they might miss out on important events that can be revealing of a team's strategy. All the different media (television, live covers, newspapers) communicate these, either by explicitly showing the numbers or through a simple visualization (C. Perin et al., 2013).

In this subchapter, four kinds of popular statistic visualization methods are described, as well as a group of other visualization methods.

#### **4.1.1 Column chart: Compared teams or players**

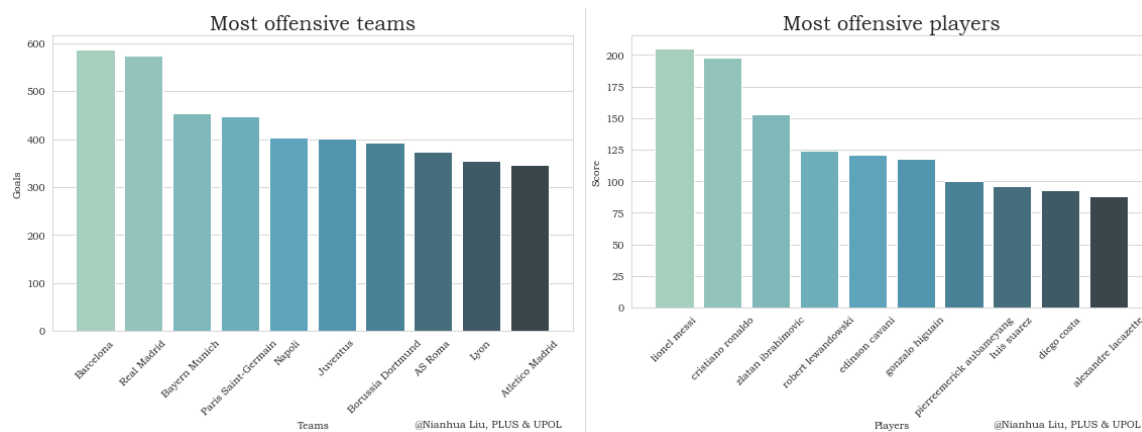
A chart is used to display data and further investigate a topic, while column chart display data in rectangular bars. The variable value of a column chart is oriented vertically, and the fixed dimension is plotted horizontally. Column chart is the most popular method in the case of comparing teams or players' statistics data. Vertical orientation of value match human cognition. Therefore, comparing the value among different teams or players is far easier on the column chart. The negative values are represented in a downward direction, and the positive values are represented in the upward direction.

To create a column chart, Excel is the most convenient and frequent tool. However, Excel relies too much on manual operation in data pre-processing and cleaning work. Besides, Excel cannot complete all the workflow independently. In this thesis, standardized and reproducible workflow is one of the considerable factors. Therefore, Python is used as the first step of the workflow to collect the data from API, formalized a unified format, process with specific index, and visualize them.

During the visualization of the data, care was taken to keep the same colour of the same variables, so the reader can focus on comparing the data without distraction to recognize the colours. The number of colours should less than seven, ideally less than four. The tone of colour will also affect the viewer's experience. For example, red bars effective to depict negative earnings. For the fonts in title and labels, Bold or Italic style should avoid appearing simultaneously, which is hard to read. A suitable size of font is necessary for viewers. Both too big and too small should be avoided. Capitalizing all the letter is meaningless, which will make the chart too condensed to read (Lacome et al., 2018). Try to make a balance between information density and aesthetics. The information load should be considered as a property quantifying the amount of content in visualization products. Graphic Map Load Measuring Tool can be used to evaluate the complexity level (Barvir & Vit, 2021).

In Jupyter Notebook, seaborn, matplotlib, plotly, numpy, pylab, and pandas libraries are imported first. With the rcParams and fontManager function, Bookman Old Style is set as the font globally. The dataset is loaded, and the needed features are selected. The main dataset was fetched from Kaggle Football Events (Kaggle, 2017). The data provides a granular view of 9,074 games, totalling 941,009 events from the biggest 5 European

football leagues. Next, the type of events and leagues are arrayed as series. After that, all the data have been prepared. In the next step, a utility function to plot bar plots with similar configuration is defined. Inside this function, figure size, style, and label size, colour, location can all be customized. Data can also be plotted in through this function. Here, the dataframe is filtered to extract the most offensive teams and players by grouping and sorting. Finally, the chart is plotting and indicates that Football Club (FC) Barcelona is the most offensive team and Lionel Messi is the most offensive player (see Fig. 2).



**Figure 2** Most offensive teams and players

#### 4.1.2 Bar chart: Longer labels for comparison

Both the Bar and the Column charts are the most commonly used charts for tracking changes over time between different groups using rectangular bars. The longer the bar, the larger the value. The difference lies in orientation. A bar chart presents the variable value horizontally and the fixed dimension vertically. Bar charts are better for representing data with long labels, comparing to column charts. The vertical dimension gives enough space for each bar to be labelled appropriately on the category axis. In cases where the data to be represented are a lot and the labels are lengthy, bar charts are better used. Bar chart and bar diagram are frequently used in football research. Kotzbek and Kainz (2017) used a bar diagram to prove that the player's overall passing quality in the second half was far better in first half. Bar chart was also used to compare players performance in previous research (Duch et al., 2010; C. Perin et al., 2013).

Here is an example with the same Kaggle Football Event dataset used in the column chart (Kaggle, 2017). Football events in Europe's top 5 leagues will be analysed to come up with a few insights regarding football teams (see Fig. 3). The core concept of this bar chart is to clarify that which teams are the best or worst in their home and away games? All the football team statistic from 2011 to 2017 are imported. According to the research conducted by Li and Zhao (2021), average goal difference is a significant metric to capture both the inclination to win and the scoring performance. Thus, this is a good metric to clarify the team's performance among the Europe's top 5 leagues. Home and away game are separated. Each of the metric list top 10 football clubs among Europe.

The result bar chart shows that Football Club Barcelona (FCB) has the highest positive average goal difference from 2011 to 2017, while they are the second highest as away team.



Figure 3 Average team goal difference

### 4.1.3 Line plot: Tracking rank change with Bump Chart

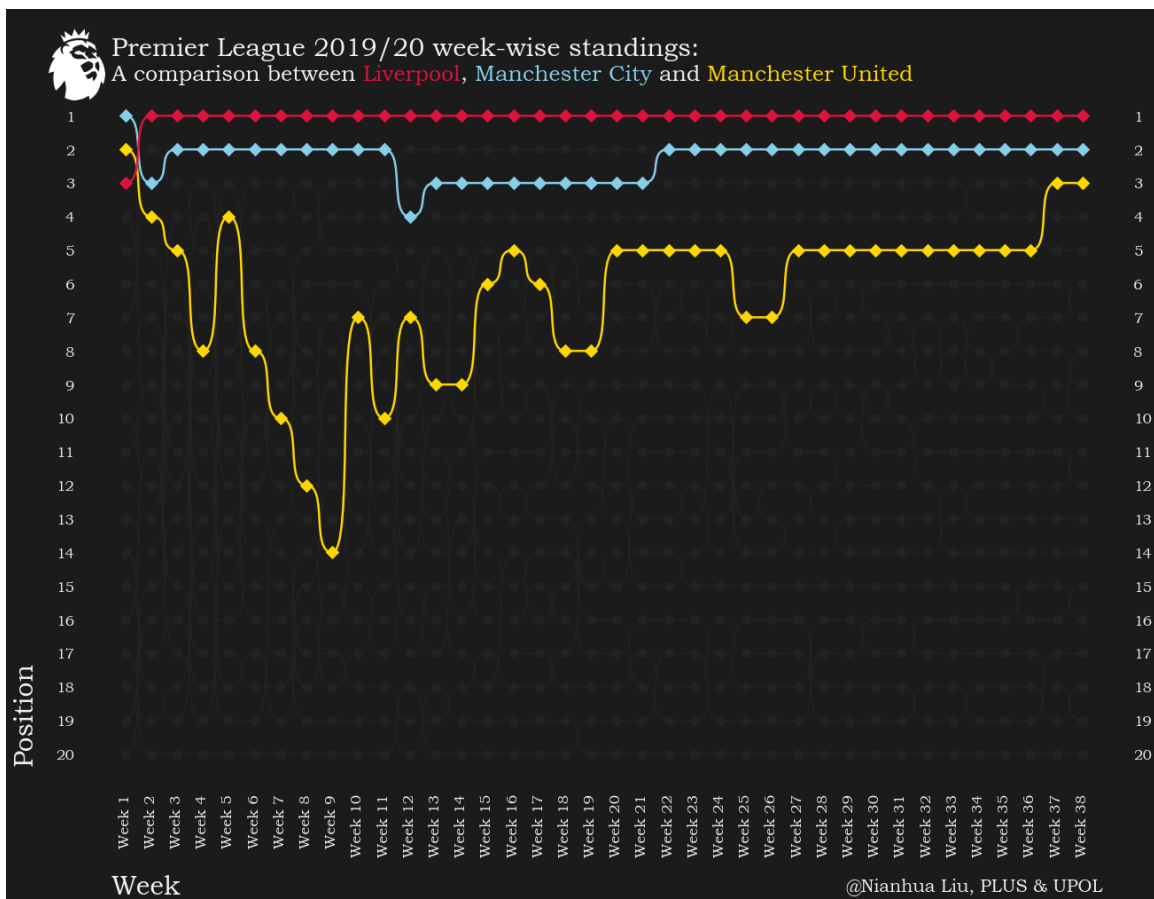
A line plot is a type of graph that displays information as points or check marks above a number line. A series of data points connected by straight line segments to plot this number line. The frequency of each value is shown through it. A line chart is often used to visualize a trend in data over intervals of time so that the line is often drawn chronologically. In football sector, line plot could be used to observe trends in time (C. Perin et al., 2013). Cotta et al. (2011) used line plot to compare the zone-based highest degree centrality among players over time.

Bump Chart is a special form of a line plot. This kind of chart is designed for exploring changes in rank over time. Users can easily track the performance, visualize changes, compare the rankings of multiple observations with respect to each other. The focus here is usually comparing the position rather than the actual values itself. When a line crosses another line, that is indicative of a change in rank. In other words, a crisscross in a bump chart indicates one football team has surpassed other in ranking. It is very effective to understand how each of the football clubs perform against the same league.

When there is a football league, there is a ranking. Bump chart is the good option to visualize the rank change of football clubs over time. Here is an example to make use of the week-wise standing data for Premier League 2019/20. The data source from StatsBomb open data was pre-processed by Andrew Rowlinson (2020), which is packaged in a Python package called mplsoccer. A good template was made by Anmol Durgapal (2022).

First, PIL, json, seaborn, matplotlib, numpy, highlight\_text, and mplsoccer libraries are imported in Python. The first step to make a Bump chart is instantiating the objects. All the markers, labels, and scatters are set in this step. Next, the values and coordinates are given to each object. Alpha value for non-shaded lines and markers are also set. Besides, Image function from PIL library is used to insert a Premier League logo to the upper left of the Bump chart. From the grey scatter background, the changing rank of all teams are visualized with transparency. The key to make this Bump chart is to highlight the line of specific teams. Red, blue and yellow are chose to represent three teams in this chart. Highlight with Darker shade or different colour has a better emphasized effect to a focal point (Lacome et al., 2018).

It is obvious that Liverpool has a big leading advantage through the whole season (see Fig. 4). They keep the first of Premier League after the second week. Manchester United were ups and downs in the whole season. They dropped to 14th at week nine and climbed back to the third at the end of the season.



**Figure 4** Week-wise standing change of Premier League 2019/20

#### 4.1.4 Radar chart: Specifically deal with player statistic

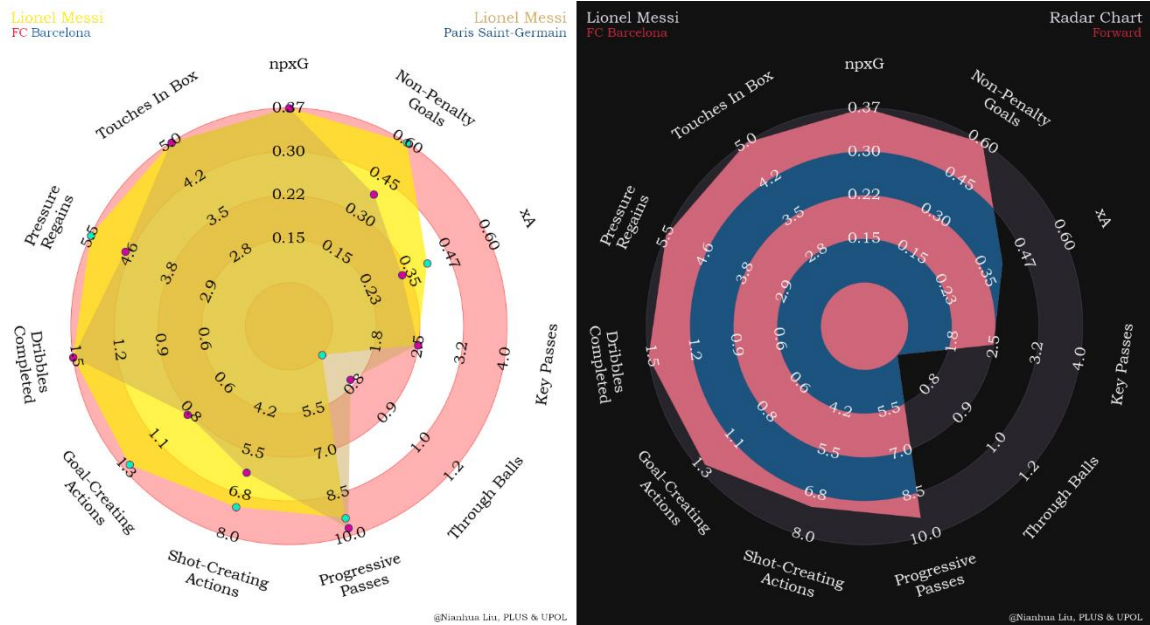
A radar chart consists of a sequence of equi-angular spokes. Each spoke represents one of the variables, while the axes of each variable start from the center. All axes are arranged radially, with equal distances between each other. The length of a spoke is proportional to the magnitude of the variable. Each variable value is plotted along its individual axis and all the variables in a dataset are connected together to form a polygon. Radar chart is a graphical method of displaying multivariate data in the form of a two-dimensional chart to compare multiple quantitative variables. This makes them useful to observe variations between them. Radar Charts are also useful for seeing which variables are scoring high or low within a dataset, making them ideal for displaying players' performance.

However, there are some major flaws with Radar Charts. First, a radar chart should not contain more than two polygons, which will make it hard to read and too cluttered. The top polygon will cover all the other polygons underneath it. Even though there are only two polygons, transparency is necessary to solve the overlap problem. Second, having too many variables will damage the viewers' experience. Seven to twelve variables are a good range after practical experience. Lacombe et al. (2018) also didn't support the radar charts with the opinion that human eyes are poor at attributing quantitative value to angles.

In the football sector, radar charts are popularly used to compare players' strengths and weaknesses. StatsBomb popularised the use of Radar for showing player statistics. In this case, the player statistics from Lionel Messi in FCB and Paris Saint-Germain are compared, provided by StatsBomb open-source data. The first step is setting the Radar Boundaries. After filtering the needed variables, the lower and upper boundaries of the bottom 5% and top 5% for the statistics should be set. Next, a radar object with the above parameters is instantiated for re-use. The radar is plotted with the setting of colour, label, title, and end note. To compare the two polygons from Lionel Messi, lighted yellow is covered in the front, and deeper yellow lies in the bottom.

The radar indicates that Messi has a better comprehensive performance in FCB than in Paris Saint-Germain, especially in Non-Penalty Goals, Goal-Creating Actions, and Shot-Creating Actions (see Fig. 5). After joining Paris Saint-Germain, Messi's position goes backward, playing like a midfielder. Thus, he has more Through Balls than the time he played in FCB.

To better highlight the statistics of Messi in FCB, a radar chart with dark these are created. Black, grey, and white become the monotone contrasting colours for background rings, titles, and labels. The statistics data is coloured by pink and blue. This radar chart better demonstrated how almighty when Messi was playing for FCB as a Forward.



**Figure 5** Lionel Messi performance visualised by radar chart

#### 4.1.5 Others: Specific purpose

Except for the most frequent charts, graphs, and plots, other statistical methods are also used in football visualization with a specific purpose. For example, Bubble chart is used to accentuate data, while Box-and-Whisker plot is used to understand distributions. Bullet chart is a good method to compare the players' performance and goals. However, some charts are not recommended in football visualization, such as 3D charts. 3D charts are fancy, but hard to interpret, which add no information to the graph in football analysis (Lacome et al., 2018).

Here is an example of Scatter plot. A scatter plot uses dots to represent various kinds of correlations between pairs of numeric variables. With a certain confidence interval, the position of each dot on each axis indicates values for an individual data point. Scatter plots' primary uses are to observe relationships between variables on the horizontal and vertical axis. Identification of correlational relationships is common with scatter plots. For example, Hughes and Franks (2005) used Scatter plot to investigate relationships between passes and goals.

From the visualization perspective, Overplotting is a notable phenomenon. When the plotting dataset is too big, overplotting can be a big issue. Data points overlap to a degree where it is difficult to observe the relationships between points and variables. The densely packed data points are hard to describe when many of them are clustered in a small area.

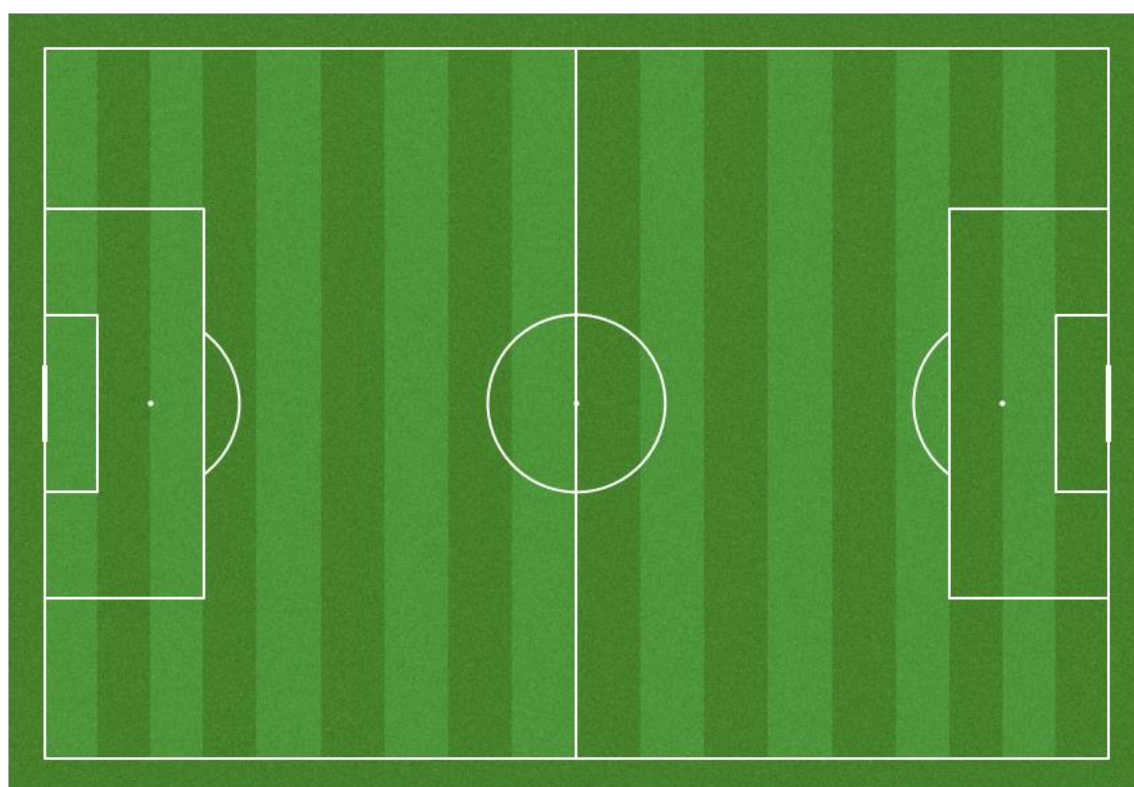
In this case, team similarity will be analysed with Kmeans clustering (see Fig. 6). The dataset is still the same one from Kaggle (Kaggle, 2017). The first step is Feature Engineering. Three features are involved here, including On-pitch violence (yellow cards, red cards, fouls), Goalscoring circumstances (play situation, play location, assist method), and Goal count. Then, three features are concatenated together into one dataset. Next, KMeans clustering is operated with sklearn.cluster library. The output data is split into 5 clusters. The best football teams in Europe are clustered on the right side. Teams from Lega Serie A are the most. Both La Liga and Bunds Liga have two teams, followed by French Ligue 1. It doesn't include Premier League clubs since at the time they weren't as strong as other clubs. They are included in the 2nd cluster to the right. Going to the left the teams become weaker.



width of the pitch, and coordinate system. The coordinate system is the first issue when drawing a match, while it is not based on any specific unit of measurement. The coordinate system varies from company to company, and is hard to unify among different data providers. Some of the local coordinate systems positioned the origin at the field's centre, while the other put it at one of the corner points.

The length and width of pitch vary from league to league too. English Premier League has an average pitch size and regulation (105×68m) for all new pitches. To better arrange the tactics, the ratio of the pitch should follow the exact dimensions for pitches in the user's league. Colours of the pitch should also follow the demand of different target groups. For example, players can better recognize their position with green grass colour background.

There are lots of existed solutions to draw a pitch map. Peter McKeever (2020) integrates all the pitch variables into a function. Ricardo Tavares (2019) and Anmol Durgapal (2022) also create some solutions to draw a pitch map step by step, from centre circle, penalty box line, six-yard box, to corner flag quadrant and the penalty spot. It's possible to draw a pitch map with only one library — `matplotlib.pyplot`. Andrew Rowlinson (2020) creates a useful package called `mplsoccer`. The orientation, colour, and axis can be adjusted conveniently through this package. Here is an example from `mplsoccer` (see Fig. 7).



@Nianhua Liu, PLUS & UPOL

**Figure 7** Basic pitch map

### 4.2.2 Positioning map

With the basic pitch map, players' location can be plotted in the positioning map. Positioning maps show the locations of one or more players at a particular time (Beernaerts et al., 2020). Positioning maps can be created by cross-sectional data or continuous panel data, depending on the analysing need. Cross-sectional data means the

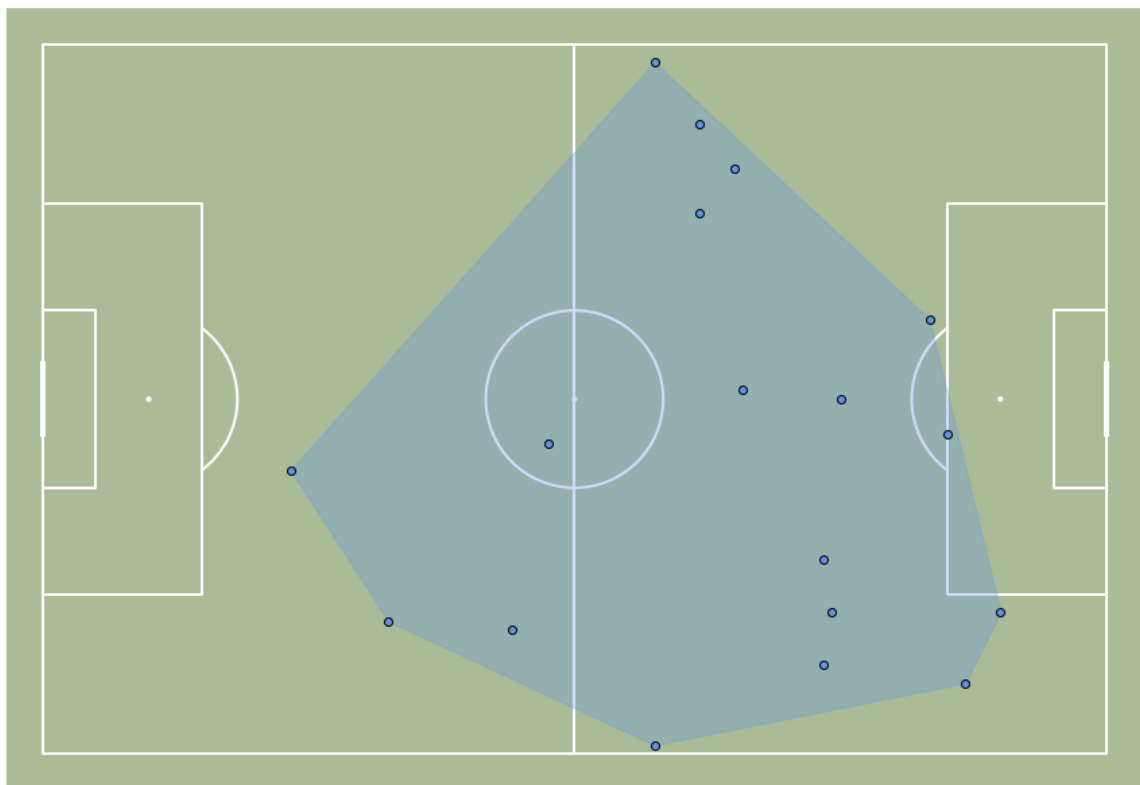


players' location at a sudden moment in a match. Continuous panel data means the moving path of the players, which are depicted as single tracking point features.

However, raw tracking data is hard to process and interpret for positioning map due to severe over-plotting and clutter when visualized. If the tracks of all players are presented, differentiation would be impossible, even for a few minutes of movement to be visualized (Sacha et al., 2017). The visual clutter of presenting too many nearby duel areas should be avoided (M. Stein et al., 2016). A generic solution to handle the trajectory overplotting problem is to simplify, abstract, or aggregate the trajectories, a concept that is known as Generalization in cartography (Harrie & Weibel, 2007). A plethora of techniques for trajectory generalization have been proposed and applied to a variety of data types and application domains (Demšar et al., 2015).

Positioning maps are widely used in previous research. For example, the scoring attempt sequences of both teams can be depicted with positioning maps (Kotzbek & Kainz, 2016). GIS tools are frequently used in positioning analysis, such as Erase, Intersect, Union, Remove Overlay, and Count Overlapping. For example, when two interaction spaces overlap, it is unclear which two players will dominate the intersection area. After buffer analysis, hatching is used to visualize these areas. Hatched areas are locations that both players can potentially reach simultaneously. In football, it is only natural that many players are close to each other and consequently several potential duel areas will arise (M. Stein et al., 2016).

In this case, a simple positioning map drawn by cross-sectional data with convex hull is shown (see Fig. 8). The data source from StatsBomb open-data was pre-processed by Andrew Rowlinson (2020). Convex hull can visualize the clusters of players (M. Stein et al., 2016). A convex hull is created based on the player's position, which could better describe the coverage area of a team.

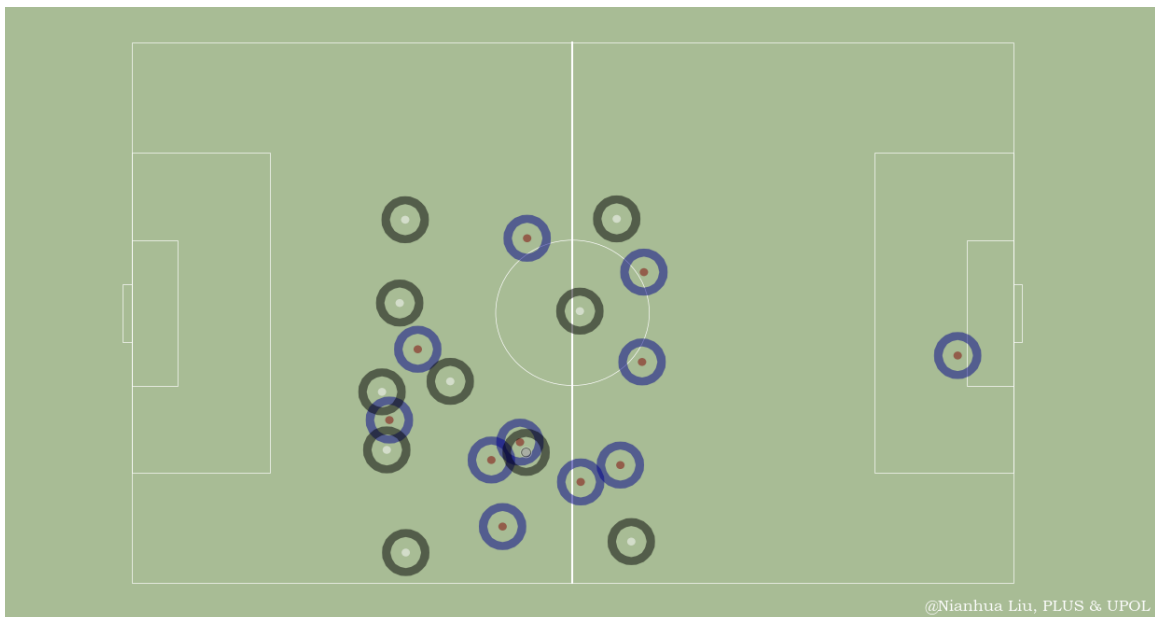


@Nianhua Liu, PLUS & UPOL

**Figure 8** Positioning map

### Buffer analysis

With the point position data, a popular GIS tool can be used in football analysis — Buffer analysis. Buffer analysis creates buffer polygon to a specified distance around the point features. The specified distance can be determined by the events, such as dribble, defence, and passes. With Buffer analysis, users can identify areas surrounding football players. This method can be used to detect players that are close to each other. For example, if the defender comes within the 3m distance buffer of dribbled player, it means that the dribble action is in danger, which needs to be careful and concentrated. Similarly, if a forward breaks through the 3m distance buffer of the defender, the goalkeeper will be in danger situation (see Fig. 9). M. Stein et al. (2016) used the aggregation-based buffer approach to depict the density of players. The radius is dependent on the size of the respective interaction space.



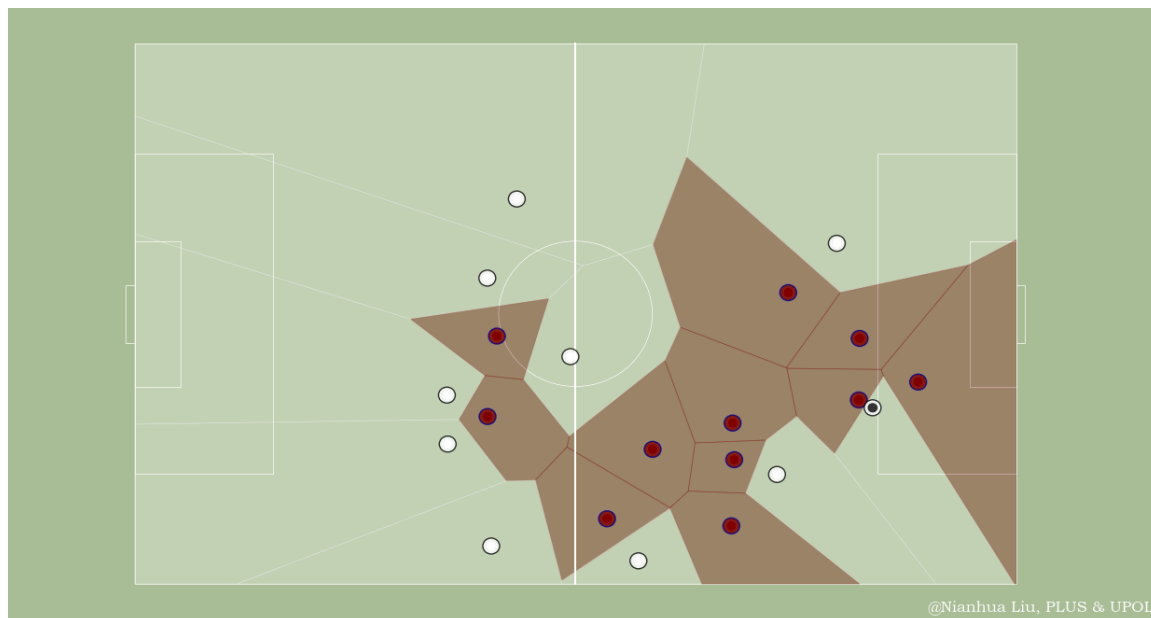
**Figure 9** Buffer analysis

### Voronoi diagram

Voronoi diagram is a spatial construction that allows a spatial partitioning of the field area into polygons. Each polygon is associated with a specific player, according to its position. A simple nearest-neighbour rule is applied to generate these polygons, which means that each player is associated with all parts of the cell that are nearer to that player than to any other player. Voronoi polygons areas can be used in evaluating the coverage of a team over the other. By introducing an excess Voronoi area, some fruitful results can be drawn to appraise a player or a team rather quantitatively (S. Kim, 2004).

Researchers have already recognized that Voronoi diagrams may provide a potent tool to study the players' spatial distribution in football. Notably, Voronoi diagrams have been applied in a variety of game settings, including real football games (Taki, Hasegawa, & Fukumura, 1996) and electronic football games (S. Kim, 2004). When real games were considered, dominant regions were calculated considering more than just players' location. In particular, Taki et al. (1996) focused on players' direction and speed, whereas (Fujimura & Sugihara, 2005) focused on the players' distance from the ball and distance to the goal. All these studies have shown that the position of the ball affects the location of the players, representing the match dynamics and hence the size of their respective dominant regions. This can be a useful method to evaluate the players' tactical behaviours during matches (Caetano et al., 2021).

In this thesis, the spatial dynamics of formation were characterized using Voronoi diagrams. Some interesting characteristics are captured. For example, players from the attacking team, tend to be dispersed whereas defenders are clustered (see Fig. 10). This spatial organization has direct influence on the dominant region defined by each player (Fonseca, Milho, Travassos, Araujo, & Lopes, 2013). These individual dominant regions were defined using Voronoi diagrams and they appear to be greater for the attacker team and smaller for the defender team.



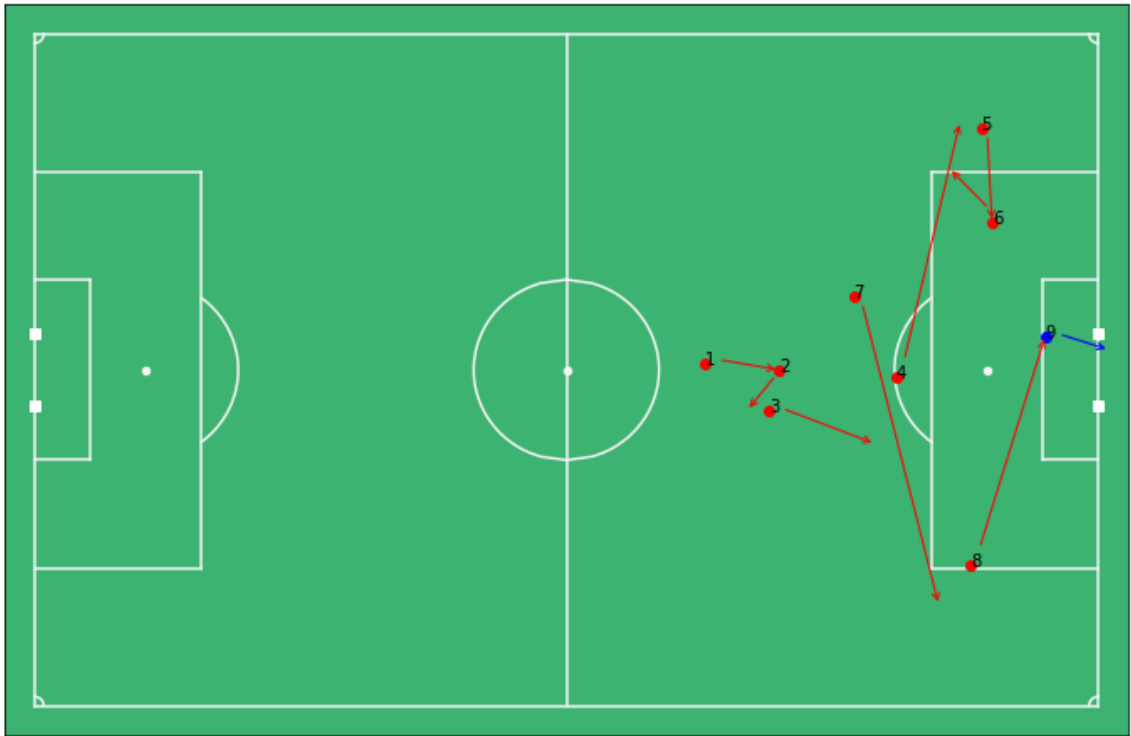
**Figure 10** Voronoi diagram

### **Movement**

For the positioning map, most of the applications used cross-sectional data. It's good to reflect the position and formation at a sudden moment. However, cross-sectional data cannot reflect the movement of players. The advantage of movement map is to depict the players' movement order and position at a certain period. It's notable that the period should not be too long, and the number of depicted players is limited. Otherwise, the trajectory overplotting problem can easily happen.

In this case, the continuous panel data is used to simulate the movement at a certain period. The key players' movement of Goal 1 is depicted (see Fig. 11). The order of his movement is directly presented for analyse. This player moves from middle to left-wing, and from left to right-wing. He finished the shooting action in the goal area, which results in a goal event.

What Happened Before Goal 1(Home Team)



@Nianhua Liu, PLUS & UPOL

**Figure 11** Players movement tracking map

### 4.2.3 Event map

The position-based specific behaviours of football players are called events, such as shots, passes, cross, tackles, offside, corners, free kicks, throw-ins, goalie kicks, penalties, etc. Notably, the event doesn't have to happen on the field, such as substitutions or cards to bench or coach team.

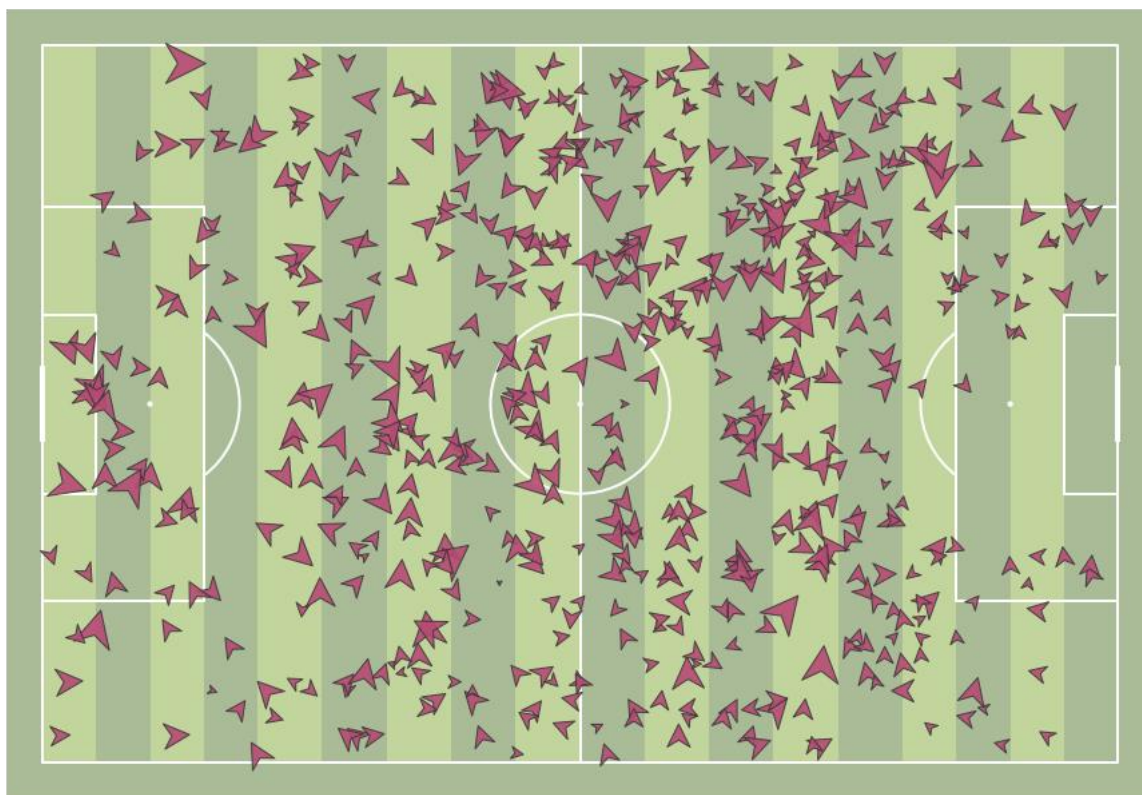
The event has to be distinguished from players' movement. Whilst the former is associated with the objects' location and spatial variance, the latter corresponds to its actions and related information (Kang, Hwang, & Li, 2006). Recording of events is different from the player trajectories in that they are not dense. Samples are only captured when an event occurs. However, event map can be semantically richer as they include details like the type of event and the players involved (Gudmundsson & Horton, 2017). While trajectories data is the primary input of the positioning map, event logs are the main input of the event map. Each row of event data should contain information about the time, space, specific players, and event characteristics. Since each point is equipped with a consecutive time stamp and frame, tracking and event data can be attributively joined.

In this case, the passes events of a football game, FCB versus Sevilla, are depicted as arrowhead markers (see Fig. 12). The data source from StatsBomb open-data was pre-processed by Andrew Rowlinson (2020). Pass event map consists of the football passing data between players, including players, passing distance, and direction. The direction of the pass is plotted to match the arrow direction. Assuming the original marker is pointing upwards, the arrow is rotated in degrees. The marker size is related to the pass distance, so larger markers mean the pass was longer. It is obvious that goalkeepers and defenders have more probability of kicking, long passes, and clear, which are longer in distance.

Meanwhile, the passes inside the attacking penalty area are much shorter. Left-wing has a greater number of passes than the right-wing.

## Barcelona Open play passes

Larger marker = greater pass distance



@Nianhua Liu, PLUS & UPOL

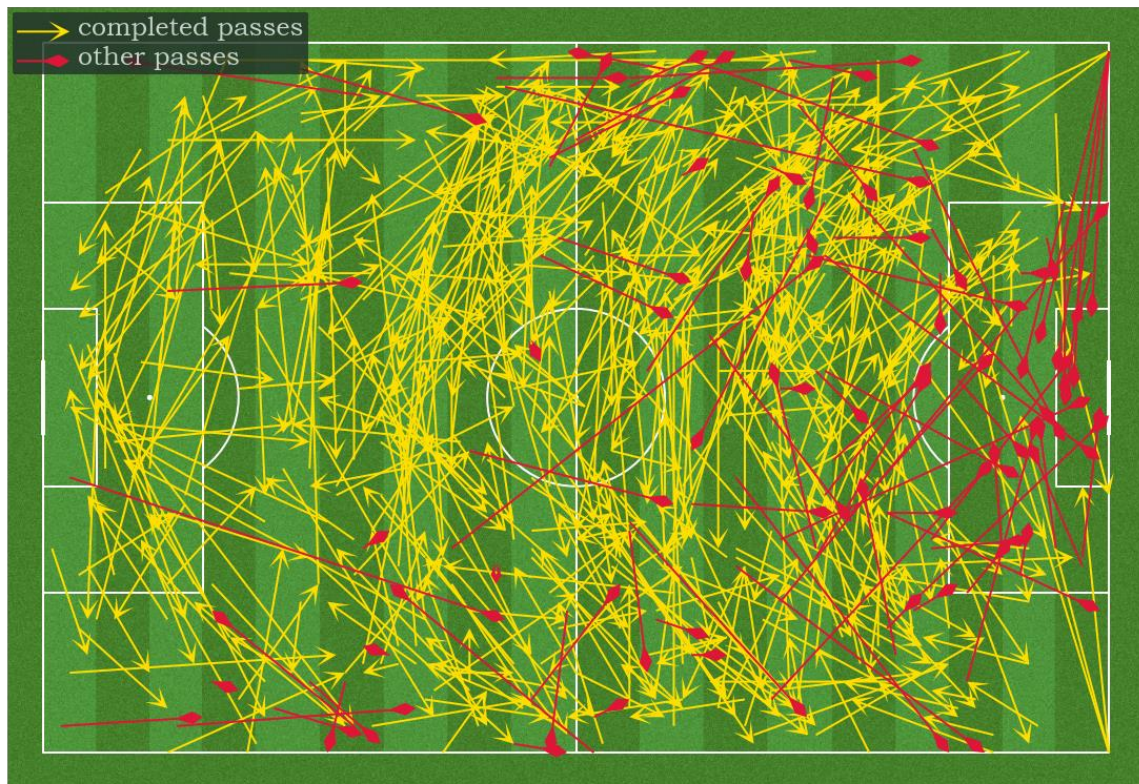
**Figure 12** Passes event scatter map

Pass event map is popularly used in football analysis. Velden, Haaren, Franses, and Hoeve (2017) create a pass map to consider the value of possessing the ball in a certain area of the pitch. With the pass event map, the researcher can find similar passes and similar attacks to value passes (Kotzbek & Kainz, 2017).

One pass could happen at a specific moment. However, there are 1000 passes in a match on average. Therefore, it's hard to plot all the passing data in a match into one map from both teams. Most of the pass maps are visualized by selected period. The length of the arrow often represents the distance of the pass.

Fig. 13 uses the same dataset as Fig. 12. The passing distance is emphasized by the length of the arrow, instead of the size of arrowhead markers in Fig. 12. This will cause the slight overplotting problem. When the lines overlap together, it's hard to separate the head of them. However, a long arrow can better capture the start and end point of a pass, compared to the arrowhead markers. Fig. 12 and Fig. 13 demonstrates that different visualization methods of event map have a big impact on the key point of delivered information.

## Barcelona passes vs Sevilla



@Nianhua Liu, PLUS & UPOL

**Figure 13** Passes event arrow map

### 4.3 Spatio-temporal Information

Football is a spatio-temporal framework with dynamic process. All football data contains spatio-temporal information and statistic information. Players movement data and events data are all collected with time stamps. Stacking the data together in a map, the players' performance over a period can be analysed. The key point emphasized by spatio-temporal information is the comprehensive performance over a certain period. Player's location at a single moment is not a factor considered in this subchapter. Their majority coverage area or regular actions, which can be depicted by time-series data, is the content in this subchapter.

#### 4.3.1 Grid and pixel map

Raw player trajectories and event data can be challenging to work with because of the low-level visualization. One way to deal with this issue is to segment the football pitch into regions with specific area and assign the location points contained in the trajectory or event data to a discretized region. When designing a spatial discretization, the number and shape of the induced regions can vary. A common approach is to subdivide the playing area into polygons of equal size. The frequency or intensity of events occurring in each region is a spatial summary of the underlying process, alternatively, the playing area may be subdivided into regions such that each region is dominated in some sense by a single player (Gudmundsson & Horton, 2017).

Grid and pixel map is a popular visualization tool when analysing the time-series data, including heat Map, intensity map, hexbin plot, and density map. The playing area is discretized into grids. If it is presented at pixel level, which is described as a pixel map.

The analysis result is distributed into specific grid. A coloured grid is drawn to visualize the results on the football pitch. The analyst can see an overview of the spatial distribution.

There are a variety of grid and pixel maps that have been used in the research. M. Stein et al. (2016) used the zonal defence map to inspect the player behaviour. Gennady Andrienko et al. (2017) used density maps to represent the spatial distributions of the outfield players. In this thesis, heat map and Hexbin plot are used as examples.

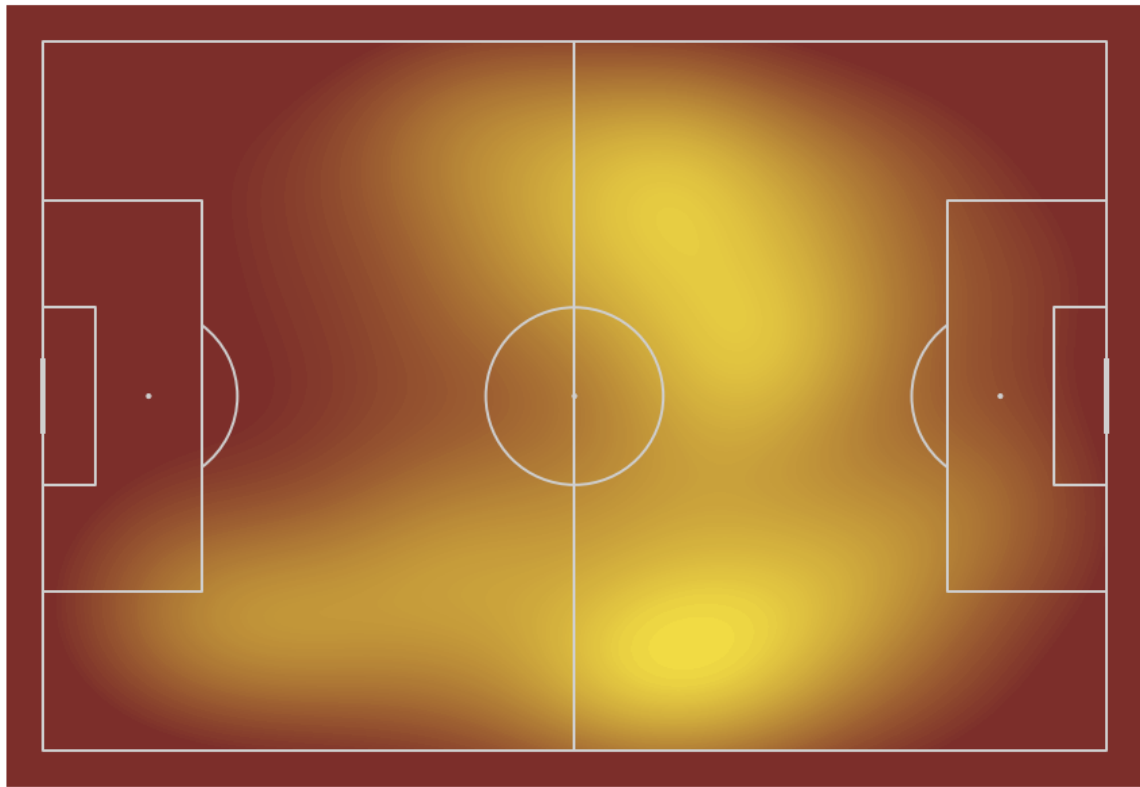
### **Heat map**

Heatmap is a kind of dynamic raster visualized symbology in the context of a football game analysis, through which player's most frequent positions are displayed by density (C. Perin et al., 2013). Heat map refers to density maps with spectral colours, which are used to identify the frequency of events that spread parts of a football pitch. Each cell in the raster visualization has a value that represents relative density. The part of the map will be heated up and turn brighter or deeper in the areas that the player can cover in a football match (Pappalardo et al., 2019).

Heat maps help the analysts know more about the players and the best positions to be placed (Pappalardo et al., 2019). The gradation of the heat map shading shows the sum of degrees of all nodes in one area. The heat map gives some information about strategies and reflects the activity of each team (Narizuka et al., 2013). Heat map could be used in drawing the control and disruption surfaces (Bojinov & Bornn, 2016) and the shots probability distribution map (Lucey, Bialkowski, Monfort, Carr, & Matthews, 2015).

The creation of color-coded heat maps must be combined with spatial-temporal data. Certainly, this kind of representation can be regarded as an improvement in respect of its consideration of space and time. However, the question of why the event happened at a particular time and place remains unacknowledged. From that perspective, heat maps have to be considered a mere eye-catcher, as their meaningfulness is finite (Kotzbek & Kainz, 2015b).

In this case, the pressure events of FCB in three matches are filtered to create the heat map in pixels, which explains how often the team was in possession of the ball in this location over the match (see Fig. 14). All teams were normalized to attack from left to right. The heat map is rendered with cmasher colormaps. On a dark background, the colormap should render from dark to light, and light to dark colormap on a light background. The tactic of FCB is obvious with heat map. Right-wing has most of the possession, especially in the midfield. This is the most frequent position when Messi was playing in FCB. Started at midfield, dribbled and sprinted into the penalty area, and finish the shooting action.



@Nianhua Liu, PLUS & UPOL

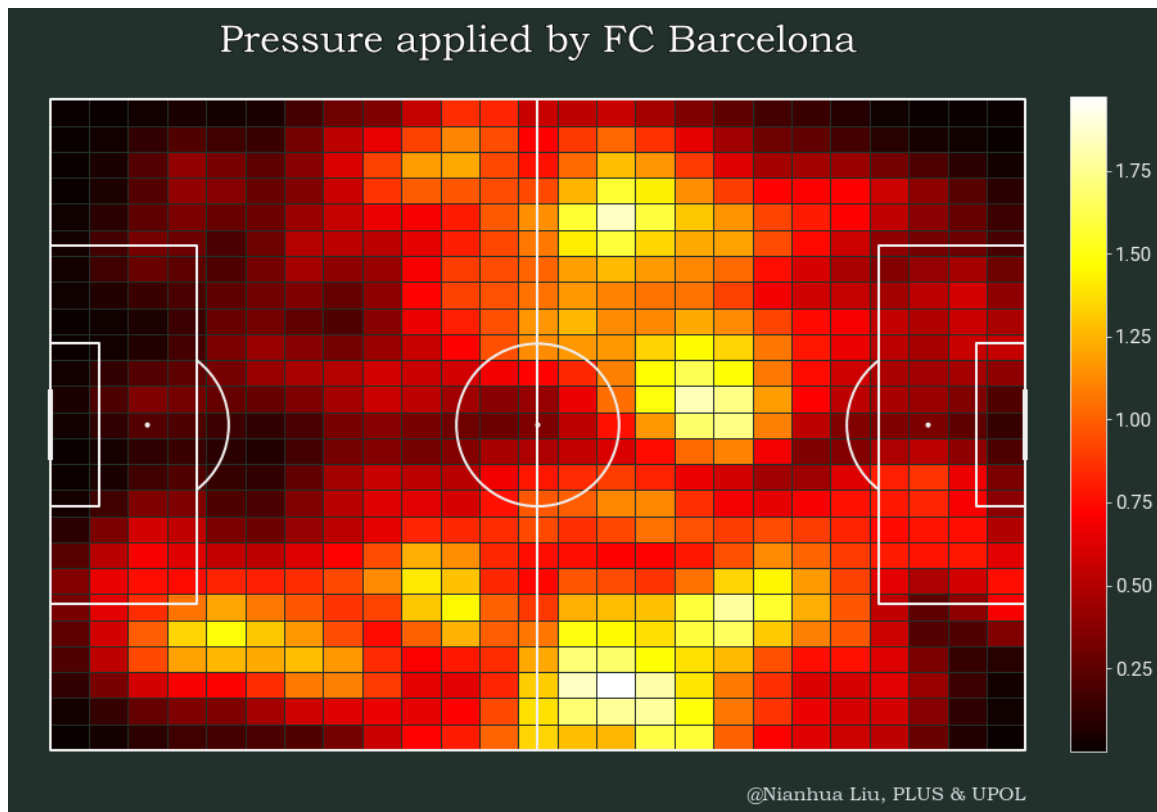
**Figure 14** Heat map

Traditional heat maps can bring the user a comprehensive perception of the density or intensity of the field. However, it's hard to locate a specific position and compare the data. A good heat map should perform more than an eye-catcher. In this case, grid map plays its role. Using grid maps can gain both a quantitative as well as a visual method of seeing occupancy of a team on the field (Lucey, Bialkowski, Carr, Foote, & Matthews, 2012).

Grid maps showing areas of the football pitch that the players occupy: Spatial data from team sports have the useful property that they are constrained to a relatively small and symmetric field. The playing area may be subdivided into regions and events occurring in each region can be counted to produce a matrix, and can be visualised with a grid map (Gudmundsson & Horton, 2017).

A visualization of a resulting ball occupancy example is shown (see Fig. 15). With the same input data as Fig. 14, the field is segmented into a  $25 \times 25$  rectangle grid, and calculated the ball occupancy of each of these grids. In this case, the frequently active area of Messi is more obvious. And this example also proves the strength of the midfield in FCB.





**Figure 15** Heat map in grids

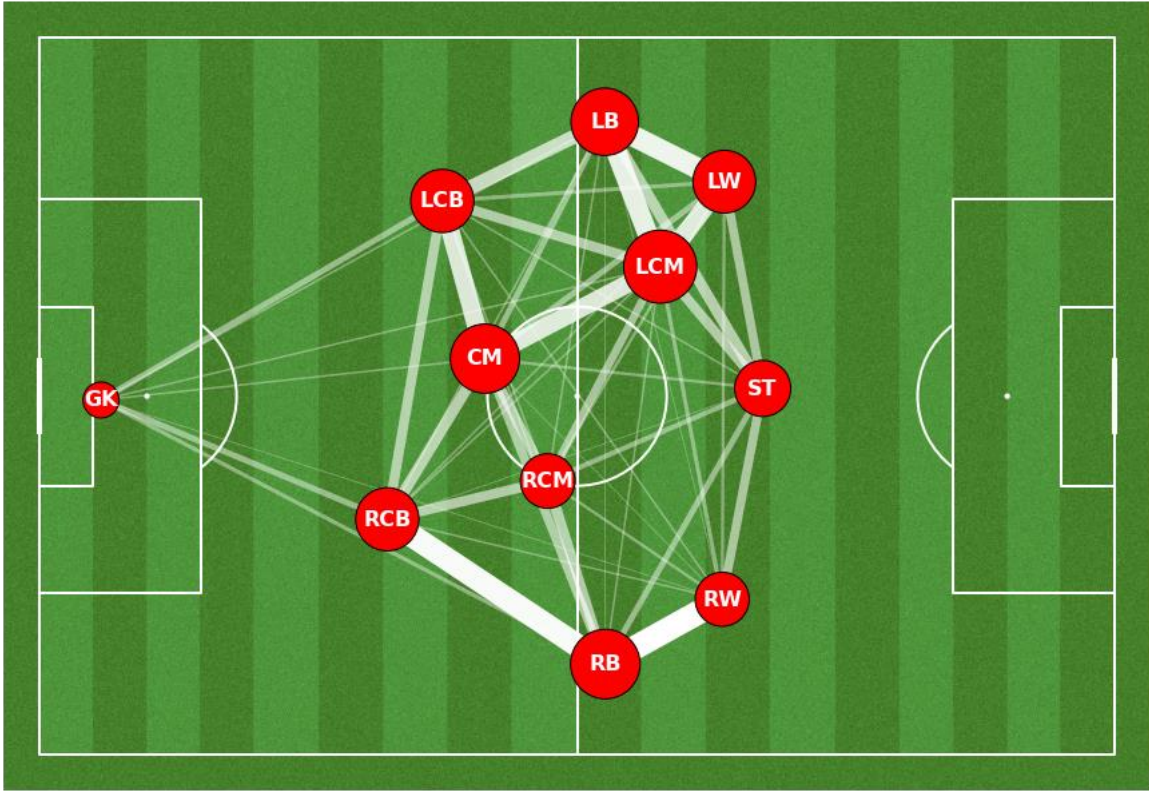
### 4.3.2 Passing network map

Understanding the interaction between players is one of the most important and complex problems in football analysis. Player interaction can give insight into a team's playing style or be used to assess the importance of individual players to the team. Capturing the interactions between individuals is a central goal of network analysis (Wasserman, 1994). An early attempt to use networks for sports analysis was in the 1977 FA Cup Final between Liverpool and Manchester United (Gould & Gatrell, 1979). Passing networks can be seen as a rich source for investigating football matches even further. Then, statistics can be extracted from a network (or graph-based) representation of the data. E.g., in ball sports, the passing network indicates which player passes the ball to which other players over time (Janetzko et al., 2014). Passing network application in football is analysed in-depth (Kotzбек & Kainz, 2017). The passing network is by all means an oversimplification of a football game, as players do not remain in static positions during games (Peña & Touchette, 2012).

In the player passing network, all players are shown as nodes and passes as links between them. Nodes are placed according to their respective player's theoretical positions. The number of nodes is constant across the different games and teams in elite football league. The position of the nodes shows player's role in the game and the links show the connections between players (C. Perin et al., 2013). If players are involved in more passes, their respective nodes have a larger size. The thickness of the network edge provides immediate insight into the interaction frequency between two teammates, with a brighter colour indicating a higher frequency of passes (C. Perin et al., 2013). The networked structure defines the passing strategy of a team during the game can be presented by the passing network (Cintia, Rinzivillo, & Pappalardo, 2015).

In this case, the passing network of FCB, in a game versus Sevilla, is depicted (see Fig. 16). Right Back is the most important position in the passing network of this game. With the connection of Right Center Back and Right Wing, the player in the Right Back has the most passes in this game.

### FC Barcelona versus Sevilla



Nianhua Liu, PLUS & UPOL

**Figure 16** Passing Network

## 5 FOOTBALL ANIMATION CASE STUDIES

Static football visualization is a good method to convey tactical instructions and present information simultaneously. However, the spatio-temporal structure decides that football is dynamic, which is hard to be visualized by the static maps. Thus, football animated visualization becomes more popular in recent years. Such measures can assist to understand the course of the game in retrospect and prepare for the tactics for the next game (Fernández & Bornn, 2018).

Animations are primarily used to depict geographic change and processes. Football animation shows the locations of all players and the ball at a particular time. Based on the additional timestamp dimension, it is possible to animate the tracking data to reconstruct the dynamic spatio-temporal structure of the game over time, either in the course of single scenarios or the complete halftime (Kotzbek & Kainz, 2015b). Animation can better help solve the over-plotting problem, while the number of players is confirmed and limited at a particular time.

However, it is notable that a long time of animation will not only make the users exhausted but also increases the total amount of data. Although every scene conveys rich information, it is too difficult to remember each frame of the animation if the length of the animation is too long. Put another way, even the amount of data that can be represented within an animation by the coach is virtually unlimited, there is a finite amount of information the football players can distil from the animation and understand the tactics output. As a result, animation is typically less than a minute in duration. They are more analogous to television commercials than feature-length films. One practical reason for this is the limitations of visual working memory (Sweller, 1988).

The trend and advantages of football animated visualization have been analysed by previous research (Fernández & Bornn, 2018). The following subchapters are each dedicated to one kind of football animation solution, based on different programming language or platform. First, the creation workflow, content, useful background information, and visualization idea is given for each of the animation topics. In addition, the main findings that can be derived for target users from viewing the animation are presented, with a short introduction of the function. In the process, noteworthy tips as well as problems and their according solutions are presented in detail. It should be noted that the dataset used in this charter is thoroughly described in the subchapter 2.1. In this charter, the introduction of the dataset will be generalised shortly.

### 5.1 Python

Python is a high-level interpreted programming language with lots of packages for all kinds of disciplines. A large pool of third-party packages about football analysis is shared in Github, which provides a good reference for this thesis. The Python animated solution was first created by Samira Kumar (2019), mainly based on the Bokeh library. Six different libraries and modules are used in this animation work, including Pandas, NumPy, SciPy, and so on. Bokeh is the most important one among them.

Bokeh is a Python library for creating interactive visualizations for modern web browsers. Beautiful graphics and smooth interface for animation are built with streaming datasets. With Bokeh, users can create powered visualizations powered by JavaScript without writing any JavaScript script.

For the current study, Jupyter Notebook with Python 3 kernel is installed on the local system. Jupyter Notebook is an interactive application founded by open-source project —

Project Jupyter. Jupyter Notebook allows data scientists and scholars to create, share, and visualize code and data. It runs in the browser and provides a simple lightweight yet powerful platform for scripting.

### 5.1.1 Creation: Visualized with Bokeh

With rich libraries, Python is a suitable solution for in-depth analysis. Thus, this subchapter aims at creating tactics visualization tools including Convex hull, Voronoi diagram, and Closest defender network. An animation slider with corresponding buttons is created to plot 2D coordinates of player movement in a game over time. For a better visualization effect, all the tactics tools are visualized with a toggle button, to show or hide the effect (Samira Kumar, 2019).

#### Pre-processing

The first step of the workflow is data collection and pre-processing. Pandas library is imported to work in this step. The data used in this subchapter comes from STATS. STATS data doesn't have player or team ID for tags. Therefore, the raw data should be pre-processed by adding tags manually. Two visualized teams are listed as digital number one and two, respectively. Player ID ranges from one to eleven for both teams. ID value 12 will be the ball's tracking data. The game sequence time is counted by seconds as the unit. A variable has been created to initialise time and calculated initial player coordinates based on that. Since the slider time will always start at zero, when the plot is displayed on browser, the time has been initialised as zero here.

Next, all the necessary packages are imported, including NumPy, Bokeh, SciPy, and Functools. NumPy is a library to add support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. Here, feature vstack is used to create the arrays, and feature where is used to return elements depending on condition. Lots of other mathematical features are used here such as sqrt, round, and abs. The Convex hull is created by the spatial module inside the SciPy library. Functools library is imported for the partial functions. Partial functions allow fixing certain values set of a function and generate a partial function application. Some portion or keywords of a function's arguments are frozen resulting in a new object with a simplified signature. In this example, make\_plot function is created based on 12 parameters (see table 1).

**Table 1** Software products and their manufacturers

Parameter	Description
Doc	Plots the graph
Df	Gets the user defined dataframe
headers	Give the headers to the dataframe - Headers should be ["x", "y", "team_id", "player_id", "time"]
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
Sport	Provide the sport details to change slider function

anim_speed	Provide speed of animation - milliseconds
show_dist_speed	Turn on/off plotting speed and distance. Default value is False.

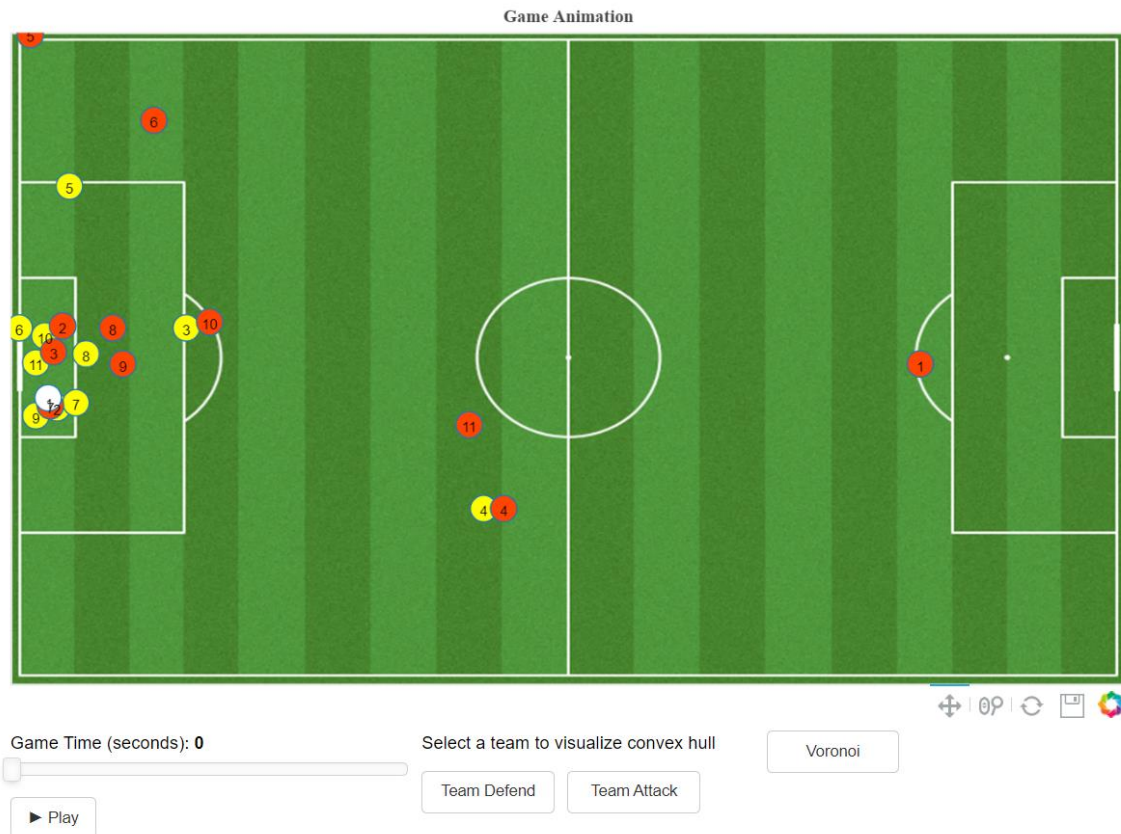
### Interface

When the data and functions are all prepared, it's time to plot the data. The method to create a pitch background is plot with an image. Here, the existed pitch map has been inserted into the specific folder (see Fig. 7). The coordinate, anchor, length, and width of the background are given based on STATS definition of the dataset. To plot the player labels and team colour, two separate variables are created. The value of player's labels and the HTML colour code of the teams are assigned to the variables with hard coding. Team one is plotted in yellow colour, and team two is plotted in colour. The ball is plotted in white colour. With the players' coordinates information, the scatter plot and label of each player can be plotted into the pitch map. To increase the robustness of this function, lots of Value Errors condition script are added. As soon as this step is finished, an individual scene of the positioning map is plotted.

Now comes the most important part — animation. In this animation, the bokeh server processing method is utilized. Over the time slots, every scene of the pitch map is altered according to the specific players' and ball's location at that moment. This technique is used to create a slider. This way allows the viewer to add widgets, trigger real-time Python callbacks, and obtain an interactive plot in the browser. Here, the io, layouts, models, and plotting modules are imported with lots of submodules such as curdoc, row, widgetbox, column, ColumnDataSource, Label Set, CustomJS, Title, Slider, Paragraph, Button, and CheckboxButtonGroup.

First, the plot will only show the player coordinates at time zero. A slider needs to add and vary the time to show the full sequence of events. To achieve this, the start and end time is defined. Next, add the slider to the plot layout using the Widgetbox. Once the slider is created, a function can be built to get the call back when slider is updated. Every time the source gets updated, all the associated plots will get updated as well, which includes the scatter plot and labels.

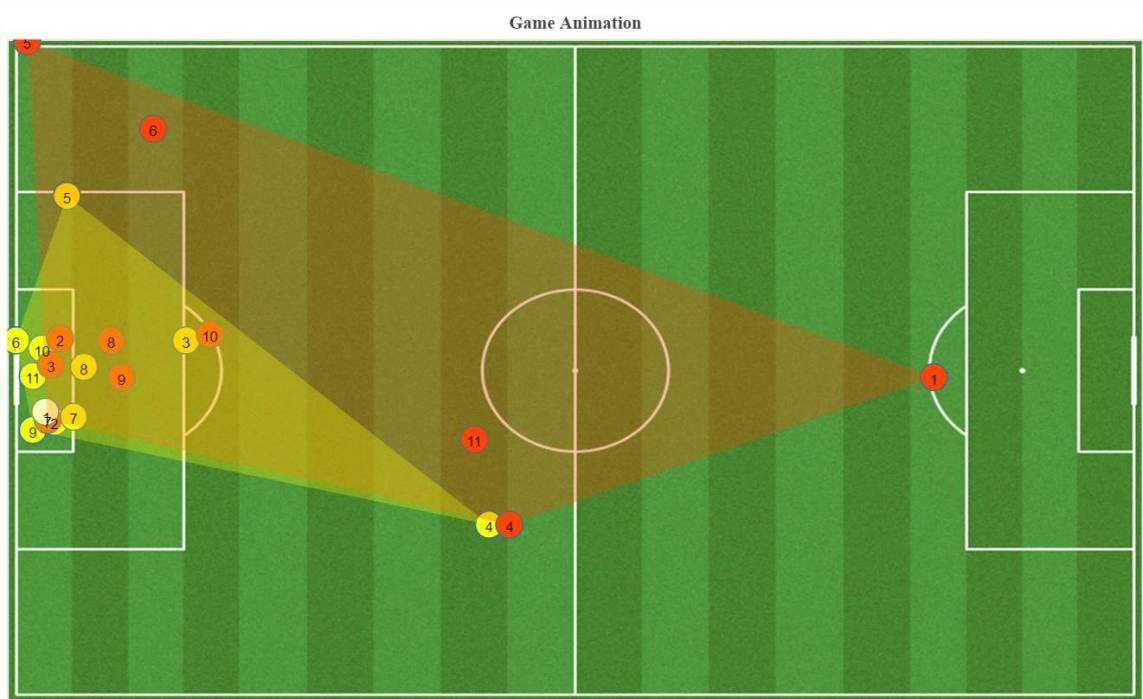
For tactics instruction purposes, the button to create the play and pause animation are created. Notably, to allow ► symbol to run in the browser, "# -- coding: utf-8 --" should be added at the beginning of the code. The animate button was first created by James Raphael Tiovalen (2022). Update the WidgetBox with the button, run the bokeh server and drag the slider, the animation will be shown (see Fig. 17).



**Figure 17** The initial interface of positioning animation

### Convex hull

To plot the convex hull, a few more variables need to be created. First, the convex hull should be created for attacking and defending teams respectively. Thus, the data frame is split to hold separate data for both teams. Numpy's `vstack` feature is used to create the two arrays. Once the arrays are created, the SciPy convex hull package is applied to get the vertices of the convex hull. Next, create two separate `ColumnDataSource` to hold these vertices values. Run the bokeh server when this step is finished, the convex hull will be plotted for the time zero (see Fig. 18).



**Figure 18** The initial interface of Convex hull animation

To create the animation effect, two `ColumnDataSource` needs to be added in the callback of the slider. In the function `update_data`, again add the above steps to plot the convex hull dynamically, based on the slider value. So now, if the viewer runs the bokeh server, the convex hull also will be updated when the slider gets updated.

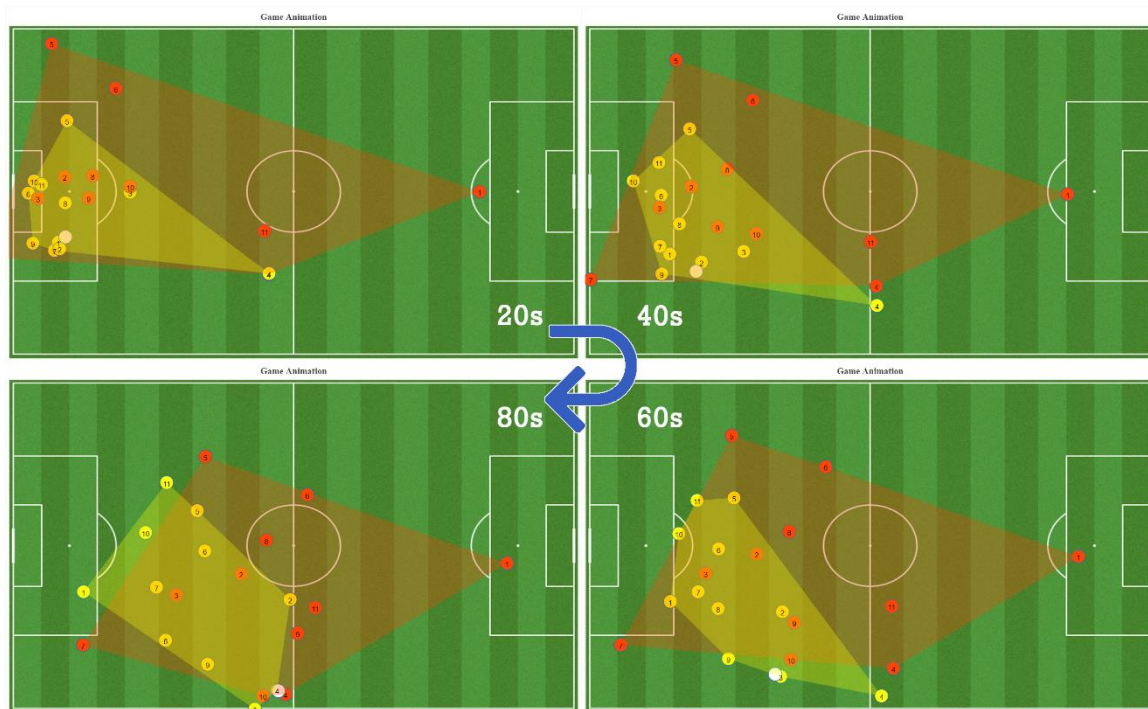
Same as the positioning animation, the button to toggle the convex hull animation on/off is created for tactics instruction purposes. A `CheckboxButtonGroup` is added to the plot. There are multiple ways to toggle the convex hull on/off. Here, the `CustomJS` option is used to set the convex hull patch alpha when the toggle button is clicked. Once the `plot.patch` is updated, the patch is not visible in the plot.

Next, add a `customJS` callback to these patches. This callback will update the patch alpha to 0.3. Add a paragraph label and update the `WidgetBox` and run the server. To create an example, four individual moments in 20s, 40s, 60s, and 80s are plotted in this thesis (see Fig. 19).

It is obvious that the attacking team, which is the red team in this example, has a dispersed formation. The players in the attacking team tend to pull apart the formation to make use of the space for open football. If the defender follows their route, the distance between the defender will be increased, which might result in losing their position. The forward will make use of this opportunity to enter the penalty area and threaten the keeper.

At the same time, the defending team has a compact formation. When the ball is inside or nearby the defending penalty area, the back and midfield will cluster together tightly to protect the penalty area. If the defending formation is tight enough, players from the attacking team cannot pass the through ball crossing the defending team.

Therefore, the coverage area of the convex hull from the attacking team is always bigger than the defending team.



**Figure 19** Example of Convex hull animation

### **Voronoi diagram**

A Voronoi diagram divides the plane into separate regions. Each region is associated with a particular point, which is the player's location in this study. The generated region indicates the area that is closer to that player's location than any other player. The animation creating process of the Voronoi diagram is similar to the Convex hull.

The Voronoi diagram is created based on the Voronoi submodule from SciPy library. The Voronoi diagram is computed using the Qhull library. Coordinates of points to construct a Voronoi diagram are the only required parameters. The furthest-site and incremental are both optional parameters.

To generate a Voronoi diagram, the first step is to input the players' coordinate point list as the parameters of `scipy.spatial.Voronoi` class. The output parameters will be the input of a function called `patches_from_voronoi`, which is used to return the coordinate values to plot patches and boundary lines. A Voronoi diagram contains two parts — the actual patches and the line segments that define the indefinite boundary. The patch is created by the vertices. A vertex is any point where lines in the diagram meet. After the calculation of centre-point, mid-point, and distant points, the boundary can be drawn. Removing patches and boundary lines that are duplicated, four parameters including the coordinate points of the patch, and the coordinate points of the line segments will be exported. Plot them in the main function and the Voronoi diagram is successfully created.

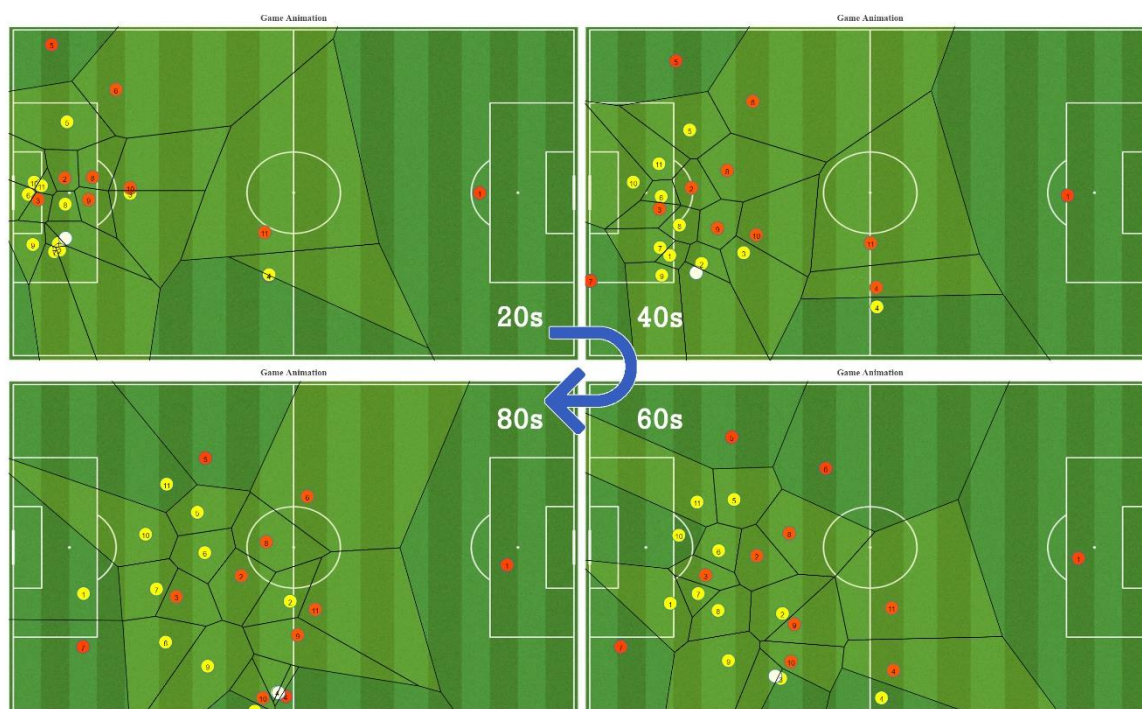
The animation creating process is the same as Convex hull. Using the `update_data` function to render a new figure every time slider is updated. Four individual moments with twenty seconds intervals are plotted as an example (see Fig. 20).

Voronoi diagram is particularly used for predicting or simulating interactions between neighbouring structures. In the football studying case specific, the position of the ball has a great impact on the Voronoi diagram. Players tend to cluster together and compete around the football. The key value of Voronoi diagram is to clarify the responsibility and collaboration of defender. The coverage area of each defender is clear in Voronoi diagram.



Voronoi diagram could help the coach to train the players on how to build a completed formation, which will not leave any space for the attacking players.

For example, the yellow team has a compact formation in the goal area in 20s. Player No.3 from red team cannot connect to any other players in the same team, which is observed from his Voronoi diagram. However, the defending strategy of yellow team for players No. 2, No. 9, and No 8 is a failure. If the ball is passed to one of them, they can easily connect to each other and make the final shot. Therefore, one of the players from the yellow team should stand in the middle of them to break the adjacency of their Voronoi diagram and cut the connection between them. The same phenomenon happened in 60s and 80s, player No. 6 in the yellow team should stand closer to the connection line between player No.3 and No.2. The adjoining line between their Voronoi diagram should be cut, which is the responsibility of player No.6.



**Figure 20** Example of Voronoi diagram animation

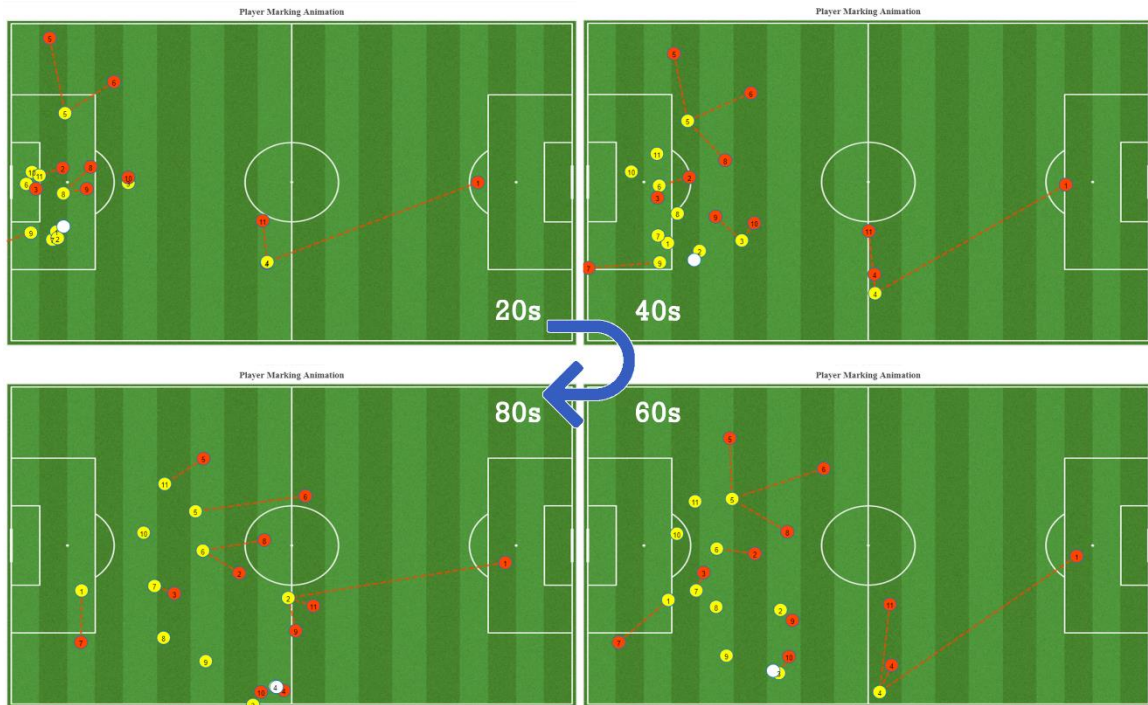
### Marking

Marking is a popular defensive strategy in football where defenders are assigned a specific opposition player to mark rather than covering an area of the pitch. While the Voronoi diagram indicates the coverage area of each player, marking is more direct to show the connected line of attacked and defender.

The principle of creating a basic marking graph is simple — calculating the nearest defender for each given attacker and making a connected line between them. The calculation is completed based on the `scipy.spatial.distance.cdist` submodule. The distance between each pair of the two collections of inputs is computed. First, the coordinate array is input in the function to calculate the closest pairs of each attacker. The output is renamed and merged to create the plot. The animation process is the same as the above two applications.

In this thesis, a phenomenon can be observed (see Fig. 21). There is a higher probability that more-to-one marking will appear when individual attacker comes close to the own half or penalty area of defender. If the player is far away from the own half, there is no necessity for marking. Thus, some of the players in opposing half have no

marking players. For example, player No. 10 and No. 11 have no marking players in 40s and 60s. Marking is a good visualization method for tactics instruction. Combined with the Voronoi diagram, the coach can better explain the zonal or man-marking strategy.



**Figure 21** Example of marking animation

### 5.1.2 Function: In-depth Analysis

Due to the large pool of third-party packages in football analysis, Python is a perfect tool for in-depth analysis. Several spatio-temporal analysis tools can be integrated together for different tactics purposes. With the Bokeh library, the animation creating process is much easier. Unfortunately, Bokeh library has a significant drawback. The standard export format is Scalable Vector Graphics (SVG) and Portable Network Graphics (PNG). Video and animation format, such as MP4 and GIF, is not the standard format of Bokeh. Thus, making a sharable animation outside the Python environment is not easy work.

Furthermore, Python needed to be executed in a specific environment. Even though it is high-level interpreted programming language, the learning curve of Python is still too sharp for sports professionals. To modify the code and maintain the application will be tough work without a programmer. Besides, Python is a language designed for Personal Computer (PC) environment in coding and executing work, instead of the mobile devices like laptops or Tablet. It's not convenient to input the data, modify the code, and execute the application when the coach team is sitting on the sidelines.

Therefore, the target user for Python solution is the football analyst in pre-game and post-game user scenarios. When a football match is over, post-game analysts collect the data, filter the specific period, complete the visualization work, and deliver it to the coach team. Before the next match begins, pre-game analysts make the new tactics visualization work according to the demand of coach team. Real-time tactics analysis is not a suitable application scenario for Python solutions.

## 5.2 Web Technology

Python is a good programming language for analysis. However, visualization is not its strength. In real-time football analysis, conveying the information with fantastic visualization is not the target. The target is to deliver the information in a fast, simple, and proper way based on the user's demand. Following this target, one way to do this is an interactive presentation where it is possible to show additional information by interacting with each data entry. Previous research has demonstrated how football statistics can be improved by interacting with different datasets (Albinsson & Andersson, 2008).

In this case, the solution should be built based on a direct interaction methodology. The data input and visualization output function should be integrated on the same page. The input parameters should be clear to connect the frontend to the backend. Web technology is a good solution to realize this such as the standard markup web programming language — HTML5. For a long time, web graphic rendering was only feasible using third-party plugins such as Flash or Silverlight. With the development of HTML5, some functionalities can be achieved without the need for any plugins, not only for the graphics matter itself but also for the audio and video.

However, HTML5 on its own cannot be used for animation or interactivity. It must be supplemented with JavaScript or CSS. The whole web technology solution will be delivered as web interfaces structured in HTML5, styled with CSS, and its behaviour controlled by JavaScript.

In fact, JavaScript probably plays the most important role, since it is the technology that implements most of the interactive features of the web pages that will be developed. JavaScript is a lightweight, interpreted, or just-in-time compiled programming language of the Web, which turned out to become the default browser side scripting language. Despite the bad reputation it had for almost a decade since it was created in 1995, JavaScript contains lots of useful libraries such as jQuery. jQuery is a fast, small, and feature-rich JavaScript library with a combination of versatility and extensibility. It is designed to simplify HTML DOM tree traversal and manipulation, event handling, animation, and Ajax. It makes things much simpler with an easy-to-use API that works across a multitude of browsers, which results in the most widely deployed JavaScript library by a large margin (jQuery, 2022).

CSS is a simple mechanism of a document written in a markup language such as HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript. CSS is used for describing the presentation for adding style to Web documents, such as fonts, colours, and spacing.

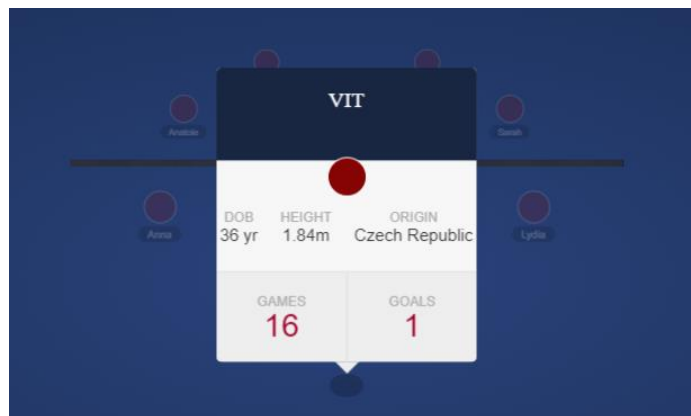
### 5.2.1 Creation: HTML5, JavaScript, and CSS

The technologies mentioned above are the foundation for the development of the football animation interface. The main interface has two different layers: an object animation layer and a data input layer. The first one aims to be the visualized layer for all football players. On the other hand, the data input layer will be the one responsible for a structure to show the information regarding a certain shape of the pitch, information of players, and tracking data of players. An example of the object animation layer is shown (see Fig. 22), where a set of players with circular shapes from home and away teams. In this case, the objective is to represent the 4-4-2 formation of the home team in red and the 4-3-3 formation of the away team with FCB icon. The Player's icon can be inserted to show in the pitch or pop-up with image format.



**Figure 22** Home and Away team initial page

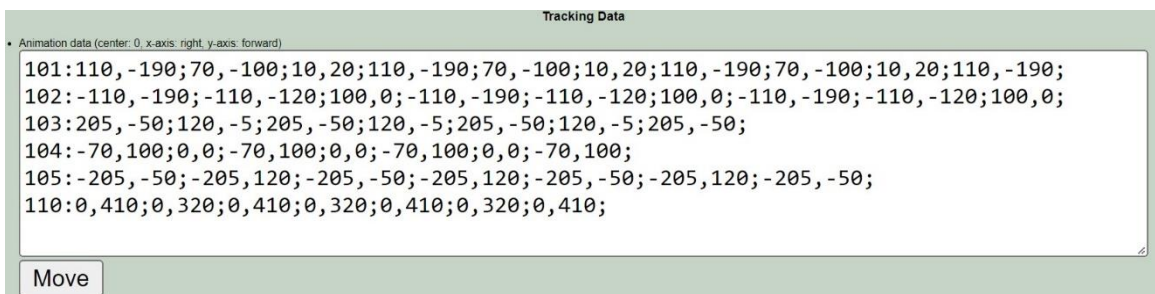
In the players' information layer, a pop-up group will be created for each player, putting together its name, age, height, nationality, and some football statistic data such as the number of matches and goals. By single-clicking on each player, players' information will pop up. Besides, icons representing the players can be inserted in image format.



**Figure 23** Pop-up information example of a player

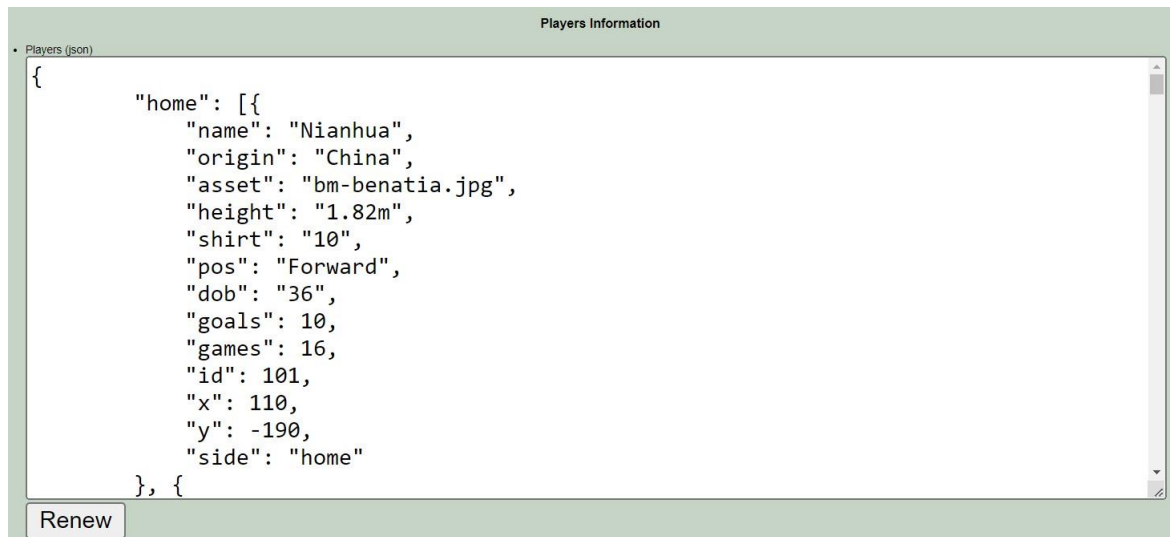
The data input layer has three parts, including tracking data of players, information of players, and pitch setting. Tracking data of players and information of players are input with JSON file format. Fig. 24 shows the example of trajectories data for each player. The origin of the coordinates is in the center. The coordinate system is defined by two axes. For the x-axis, the right side is the positive value, and left side is the negative value. For the y-axis, positive value means the player's location is closer to the viewer and bigger.

Notably, visual analysis is applied in this web visualization. The animation and the trajectories data have been transformed for the 2.5D effect. All the icons have been set in the same size initially. When the icon moves closer to the view, the size of the icon when increasing according to the distance to the screen. This is more real as a coach vision. Using the transformed trajectories, the movement of the individual players within the group can be investigated and compared quickly. To reduce the display clutter, users should not input too many trajectories at one time. One or a few trajectories should be input first and followed by more trajectories later.



**Figure 24** Tracking data input box on the initial page

Similar to tracking data, players' information is also input in a JSON file format. The parameters contain name, nationality, icon, height, number, position, age, number of goals, number of games, unique ID, default coordinates, and team. Default data and format is saved in the `configure.js` file. Players' information is connected to the pop-up (see Fig. 23).



**Figure 25** Players information input box on the initial page

The third input box is about the pitch setting, including the parameters like pitch width, pitch length, side line width, side line length, kick-off circle width, kick-off Circle length, penalty area width, penalty area length, goal area width, goal area length, penalty arc width, penalty arc length, and midfield width. The difference between the pitch setting is the input format is limited to the digital number as a static class.

Three input boxes are integrated together as a setlist. The setlist can be closed for aesthetic purposes. A tip for saving the workload of building these input boxes is creating the corresponding elements. For example, the function of the click button, text input box, and number input box are all packaged in the `utils.js` file. When it is needed, the element will be called by the HTML type in `configure.js` file.

The whole solution is built in three parts: HTML5, JavaScript, and CSS. The page structure elements like `<main>` and `<script>`, and section element like `<head>` and `<body>` are held by HTML file. The `<body>` element contains all of the content of the document. The `<main>` element is used to designate the unique and main content of the page. The `<head>` element provides a collection of metadata information for the document, including its title and definitions of scripts and style sheets. The `<script>` element is used to embed executable code or data. Notably, the id value of HTML file should be unique, which should be not repeated.

The next point to look at is the implementation of the major features with custom JavaScript code. The base architected structure using jQuery is schematized. Once again, JavaScript will play an important role to get more functionality from HTML5. In this case, some visualization effect, such as switch sides and shaking effect, is created and integrated in `Index.js` file. The connection between the player and tracking data is the `divmove_steps()` function. When the ID of a player is selected, the moving path will be matched. The path is created by points with JSON file. To renew the trajectory data, the only needed operation is to maintain the JSON file, which is convent for the users.

Next, the animation can be created based on the input trajectory data. The most important function here is the `applyEffect()`, which is used to calculate the position of the

players for every moment. The coordinate data, duration time period and fps data are the input parameters. Notably, whenever the user initialized the players, it is necessary to call the `init()` function.

Fig. 26 shows a very basic animation: the red and the blue players moving, and the switch of the pitch. For example, six players whose icon is coloured in red are observed when moving. Even if the forward has more moving distance than the keeper, the vision distance looks similar because of the 2.5D effect.



**Figure 26** Example of web animation

## 5.2.2 Function: Interaction

The main advantage of this web solution is interaction and convenience. In specific, the visualization of multiple football players is coordinated through synchronous highlighting of corresponding colours, uniform reaction to interactive actions, synchronous animation, and other standard techniques (Natalia Andrienko et al., 2013). The target user group of this solution is the coach team. As the thesis mentioned above (see subchapter 4.2.1), the coach prefers to use the vertical pitch map to emphasize the attack and defence direction. That's why this solution is designed in the vertical direction. It is possible to represent a tactical movement regardless of its complexity. This structure might suffer a few tweaks in the future, but until this date, it is the appropriate one in real-time football analysis.

For example, each football coach has his own tactics and formation, which makes it important that the input box is easy to adjust by the demand of the coaches, but they should also be as customizable as possible. A formation of 11 players can be presented by coach to players before the match start. Before the substitution, it could also change the movement guidance of individual players.

For future work in this interface, the interaction methodology can be chosen between the data input, and the drag and drop. Each of them has pros and cons. For data input, it's easier to connect with the tracking hardware and transfer the data. However, the input time might be longer. For the drag and drop, it's easy for the coach to understand and adjust. However, the data is hard to transfer without access to the backend.

## 5.3 R

R is an open-source interpreted programming language for a wide variety of statistical computing and graphics techniques. Users have created thousands of packages to augment the capabilities of the R language (R Foundation, 2022). In this case, R software and the packages are all downloaded via Comprehensive R Archive Network (CRAN) through RStudio Desktop.

RStudio provides open-source programming software for R, which integrates the development environment. It is available in two formats: RStudio Desktop is a regular desktop application while RStudio Server runs on a remote server and allows accessing RStudio using a web browser (RStudio, 2022).

### 5.3.1 Creation: Standardized with ggsoccer

In this example, five packages are used to create the animation, including ggsoccer, gganimate, ggplot2, dplyr, and lubridate. ggsoccer is a R package specialised in football visualization. Because ggsoccer is implemented as a layer of an open-source data visualization package called ggplot, it provides a set of functions that make it easy to plot and customize soccer event data in R environments (ggsoccer, 2022). ggplot2 is a R package dedicated to data visualization. It can greatly improve the quality and aesthetics of the graphics (ggplot2, 2022). gganimate extends the grammar of ggplot2 package for creating animated ggplots. It provides a range of new functionality that can be added to the plotted object and customize how it should change with time (gganimate, 2022). dplyr is a grammar of data manipulation, which is designed to enable dataframe manipulation in an intuitive, user-friendly way. Lubridate is an R package that makes it easier to work with dates and times.

Tom MacNamara (2020) first created an animation with ggsoccer package. An open-source dataset collected by Metrica has been used, including event and tracking data (Metrica Sports, 2020). Tracking data is used to plot the players on animation, and the event data can help to find the meaning period in the game to plot. Notably, different data providers may use alternative coordinates. The default setting of ggsoccer is the Opta's 100x100 coordinate system. ggsoccer also provides support for a few data providers such as Statsbomb and Wyscout.

Data from Metrica stores the value of range 0 to 1, while ggsoccer uses a scale of 0 to 100 on both axes. To transform the tracking data and match the Opta's 100x100 coordinate system, every value in the coordinate column needs to multiply 100. Some column headings are altered to make the data clearer. Besides, the coordinate data of players on the bench is marked as NaN (Not a Number), which is removed to increase the data loading speed.

Looking deeper at ggsoccer packages, five indexes are provided to customize the visualization work. Some important indexes are listed here. `annotate_pitch()` can add soccer pitch markings as a layer for use in a ggplot, such as colours of pitch line and fill. `pitch_opta()` is the coordinate system used to customise by supplying a pitch specification to the dimensions argument of `annotate_pitch()`. ggsoccer provides pitch specifications for a few popular data providers by default. However, user-defined specifications can also be used (ggsoccer, 2022). `theme_pitch()` is used to remove background and axes details from a ggplot plot.

Except for the above indexes, `geom_point()` also needed to be mentioned. Normally, `geom_point()` is used to draw a scatterplot defined by an x and y coordinate. Here, the information about the ball's location, players' location, size, and color are saved in

geom\_point(). geom\_label() is a necessary index for animation, which added a clock variable to the data to show the match time used.

In this case, a goal event happened around 90s. This is a classical corner tactics event. Setting the period from 85s to 93s, the goal event is completely revealed. The crucial player in this event is filtered to highlight their contribution and mistakes. Here, four players coloured in red are the main contributor of this goal event. Each of them is marked with a geom\_point() for plotting. A clock variable is added to the data to show the match time using geom\_label (). The lubridate package is used to get the change time for animation.

After all, animate() is called to plot the whole animation (see Fig. 27). When the short pass between the red players happened at 86 second, the yellow player cannot press the forward tightly. And then, he cannot block the crossing to the penalty area, which is the second mistake. It's clever to change position for the red players in the penalty area, which break loose of the defender. The other mistake for yellow players is that they cannot marking to the poacher to stop him dash forward and breakout. A series of contributions and mistakes make this goal event happen finally.



**Figure 27** Example of R animation

### 5.3.2 Function: Data standardization

Compared to the powerful analytic solution of Python and the convenient solution of web technologies, the biggest advantage of R solution is the standardization. As the formal description of ggsoccer package, the package is designed to work smoothly with a variety of popular sports data providers (ggsoccer, 2022). Thus, if the user hopes to visualize the tracking data from the popular sports data providers in a convenient way, R is a good solution to do so. Besides, ggsoccer only provides five indexes, which is much less than other solutions. Although this is not perfect to fulfil the customized demand, it is benefit for decreasing the threshold of football analysis.

Furthermore, R packages like ggsoccer, gganimate, and ggplot2 are interconnected with each other. The calling process of the package is not complicated. The time to build



an R solution in football analysis without any previous experience is the least comparing to Python and Web technologies. ggsoccer package provides a set of functions for elegantly displaying and exploring soccer event data with ggplot2, which makes it easy to plot data on a pitch. Simply call ggplot and use the pitch annotation and theme is enough to draw a pitch. Therefore, R is a good solution for beginner of football analysts, especially when they are using football data from a popular business data provider.

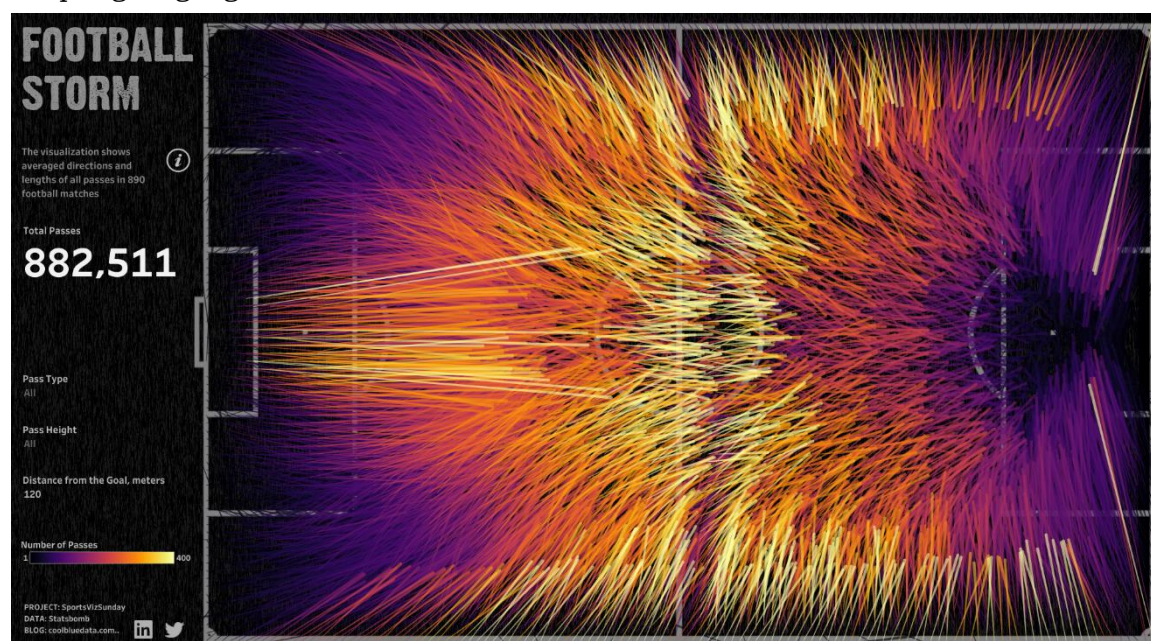
## 5.4 Others

Except for the above solutions, there are some other technical solutions to visualize and animate the football data. Each of them has specific strengths and drawbacks. Open-source, free, and popular are the main reasons that Python, Web Technology, and R, are listed as the main recommended solutions in this thesis. Here, some other solutions are simply introduced.

### Tableau

Tableau is a powerful tool with a straightforward interface to create data visualizations that can be used in a variety of settings. The public version of Tableau is free for everyone, enabling users with less programming background to produce interactive graphs. A large pool of study cases supports tons of potential uses for learning. An extensive gallery of infographics and visualizations that have been created with the public version. Alexander Varlamov (2022) has created lots of amazing Tableau football visualization. However, Tableau is not an open-source solution. Created by a business company, Tableau Creator, Explorer, and Viewer versions are expensive.

Although Tableau has a strong visualization tool, it is not designed for animation, especially in the web environment. It is feasible for Tableau to create some animations, but the animated effect and visualized channel are not friendly. The data loading speed of Tableau in a web browser environment is not fast enough to make a smooth animation. Besides, if the user hopes to make some football analysis in Tableau, such as Voronoi diagrams, they still need to load the whole position data, process it in Python, and use the User Defined Function (UDF) framework of the analytics database with the help of a scripting language docker container.



**Figure 28** Football Storm created in Tableau (Alexander, 2022)

## **ArcGIS and Adobe products**

Cartographers are familiar with Esri products, especially ArcGIS pro and ArcGIS Desktop. To enrich the visualization effect of ArcGIS, Esri developed an extension called ArcGIS Maps for Adobe Creative Cloud. ArcGIS Maps for Adobe Creative Cloud enables professionals to access geographic data in a native format and design data-driven maps inside Adobe Illustrator and Photoshop. With the extension, vector and raster map content is converted into editable, layered artwork for an improved editing experience in Adobe Illustrator. For designers, the spatial data is accessible in a familiar format to create visualizations.

The geographic content can be quickly found and added to work as editable vector layers and high-resolution images (Esri, 2022). But in football analysis and visualization, the solution combined with ArcGIS and Adobe products is not proper.

First, ArcGIS Pro is not a tool created for sports analysis, especially for the events played in a spatio-temporal limited framework. It's not necessary to collect the data in a geographic coordinate system, which only increases the complicity of data pre-processing.

Second, the creation steps of animation are too complicated and the processing chain is too long. ArcGIS products cannot connect with Adobe After Effects directly. From the original data to animation output, users should transfer the data into different software (Elbert & Rey, 2020). The tracking data is pre-processed in ArcGIS Pro initially and exported in Adobe Illustrator Exchange (AIX) file format. AIX allows users to export maps and layouts from ArcGIS Pro and use them with ArcGIS Maps for Adobe Creative Cloud extension to open files in Adobe Illustrator with organized layer structures for graphic editing workflows. However, every group of layers needs to be isolated manually. The animated objects need to be an individual layer. When the layers are prepared, all the elements can be imported into Adobe After Effects. Each layer is faded or popped individually. All the opacity key frames and Time-Reverse the Keyframes are selected to reveal the layers and extend the duration on screen. Every move of this processing chain should be operated manually without automatic workflow, which wastes too much time on the workflow. The visualization work of a single game is much more than other solutions, which is not proper for frequent football matches in a league.

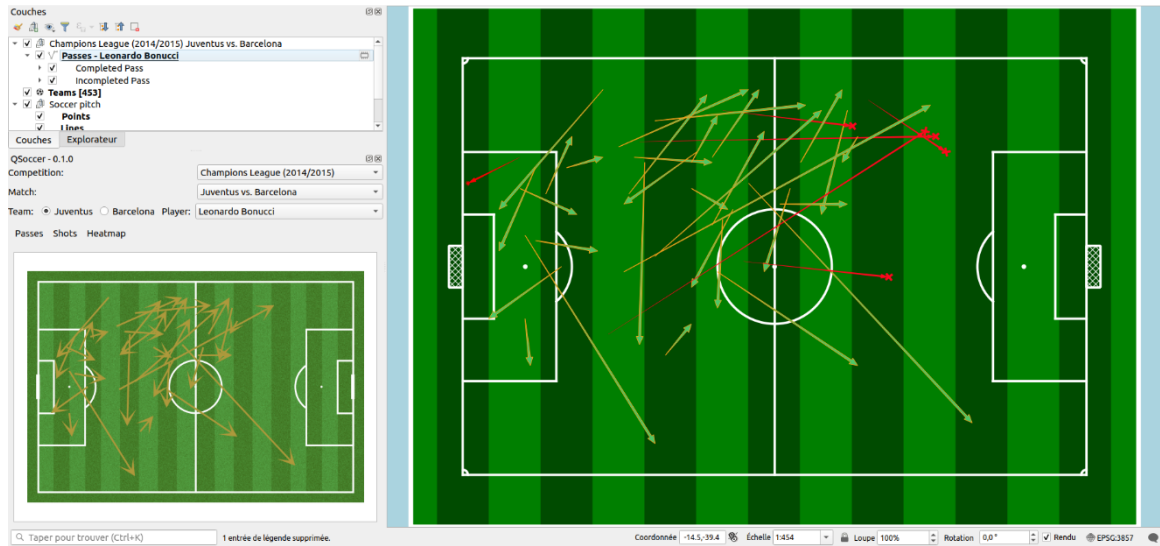
Third, since Adobe After Effects offers no native ways to import datasets of spatial file formats, a conversion of the data to an intermediate format that can be read by After Effects potentially loses essential information such as attribute tables or spatial location (Schäfer, 2021).

Fourth, ArcGIS and Adobe products are expensive. Small football clubs cannot afford the price to buy their license only for manual visualization work, without too much tactics support methodology.

## **QGIS**

After introducing ArcGIS, QGIS is inevitable. As open-source software, QGIS has a larger plug-in developer community. QSoccer is a QGIS plug-in developed by Oslandia. QSoccer aims at depicting individual player actions through a QGIS Widget. In order to localize the team on the map, one highlights the involved teams on the main canvas. The plugin support draws a pitch and depicts soccer data, such as passes, shots, and heatmap of individual players (QSoccer, 2021). QSoccer saves lots of match data from previous matches. However, as a beta version, QSoccer doesn't support importing the data with customized analysis. But still, this is a successful first step for football analysis in QGIS environment.

In the future, a potential developing direction is animating the dataset with timestamps. QGIS can visualize the data using the Temporal controller by working with a dataset that contains timestamps or has observations recorded at multiple time steps. The temporal controller allows the user to view and export 'slices' of data between certain time intervals that can be combined into animations (Animating Time Series Data, 2022).



**Figure 29** Pass map example in Qsoccer (Qsoccer, 2021)

## 6 RESULTS

The results achieved through work conducted as part of this thesis are summarized in the following chapter. The main result of this thesis comprises the accomplishments regarding the framework and workflow of visualizations and animations. A set of charts, diagrams, statistic graphs, and animations are created to clarify the user scenario and application. Then, the characteristics and value of each created static visualization are concluded shortly (see subchapter 6.1). Finally, this chapter concludes with a summary of three animations (see subchapter 6.2).

### 6.1 Static Visualization

Static football visualization represents the work presented in the chart, diagram, graph, and map. The information that can be derived from each static visualization is various. Thus, they are separated into three groups — Statistic information, Spatial information, and Spatio-temporal information. They are served for different target users and applied in different user scenarios (see Table 2).

**Table 2** Static visualization.

<b>Information</b>	<b>Characteristic</b>	<b>User Scenario</b>
Statistic	Visualized the statistical data Presented without a pitch map	Retrospective analysis
Spatial	Highlight the players' position Specific location can be observed Presented on a pitch map	Real-time analysis
Spatio-temporal	Highlight players' performance in a period Specific location cannot be observed	Mixed analysis

#### **Statistic information**

Statistic information is the most popular visualized method in football analysis. Displayed without a pitch map is the most notable feature of statistic visualization. This is also the reason that statistic visualization is not suitable for real-game analysis for coach teams. It is difficult for players to understand the tactical instruction immediately. A correct user scenario for statistics visualizations is retrospective analysis. When a football match is over, the coach can compare the statistic with previous matches and made an analysis report.

In this thesis, four kinds of statistic information visualizations are introduced, including column charts, bar charts, line plots, and radar charts. Column chart is the most popular method in the case of comparing teams or players' statistics data. The colours, font, and size will affect the efficiency of information delivery. Making a balance between information density and aesthetics is important. Bar chart is a good metric to clarify the team's performance, which is better for representing data with long labels, compared to column charts. A line chart is often used to visualize a trend in data over intervals of time so that the line is often drawn chronologically. Bump Chart is a special form of a line plot. Users can easily track the performance, visualize changes, and compare the rankings. When a line crosses another line, that is indicative of a change in rank. Radar chart is useful for seeing which variables are scoring high or low within a dataset, making them ideal for displaying players' performance.

### **Spatial information**

Spatial information is continuously produced throughout the whole match with the movement of players and football. Compared to the statistic information, the characteristic of spatial information is the presentation on a pitch map. Thus, creating a pitch map according to the user demand is the initial step of spatial information visualization. Existed solutions and basic workflow to draw a pitch map are described in this thesis. Horizon pitch map is commonly used for the audience and commentators because this is their viewing perspective of a football match. The coach team prefers to use the vertical pitch map to emphasize the attack and defence direction.

When visualizing the players' position, over-plotting is a regular problem. If the raw tracking data of all players in the whole match are presented, it is impossible to differentiate them. Selecting an important period or picking a suitable visualized method can better help solve this problem. For example, presenting the pass distance by the size of arrowhead markers is better than the length of the arrow in a pass event map.

Over the players' position, some GIS analysis can be overlaid, such as Convex hull, Buffer analysis, and Voronoi polygons. A convex hull is created based on the player's position, which could better describe the coverage area of a team and visualize the clusters of players. Buffer analysis can help the coach team to instruct the safe shielding distance or team defence work. Voronoi polygons areas can be used in evaluating the individual dominant regions of each player.

### **Spatio-temporal information**

While the spatial information highlights the players' position, spatio-temporal information emphasizes the players' performance in a period, such as the majority coverage area or regular actions. Player's location at a sudden moment cannot be recognised.

Heat map and pass network are two examples presented in this thesis. Heat map refers to density maps with spectral colours, which are used to identify the frequency of events spread parts of a football pitch. Each cell in the raster visualization has a value that represents relative density. Pixel-based heat map is hard to locate a specific position and compare the data. Using grid maps can gain both a quantitative as well as a visual method of seeing occupancy of a team on the field. Pass network can better describe the pass behaviour and performance of a team. If players are involved in more passes, their respective nodes have a larger size. The thickness of the network edge provides an immediate insight into the interaction frequency between two teammates, with a brighter colour indicating a higher frequency of passes.

## **6.2 Animation**

The spatio-temporal structure decides that football is dynamic, which is somehow hard to be visualized by the static maps. To bridge the gap between the football coaches and the effect information delivery of football, three animated solutions presented in this chapter are created (see Table 3). Although the solutions are implemented in multiple languages, the strength and user scenario of each solution are depicted, which can expand our horizons in developing a cross-browser, cross-platform, and interactive application. There is no doubt that Python provides a strong football analysis skill due to its rich libraries. Web technology highlights the convenience of interaction. R can work smoothly with a variety of popular sports data providers and save the workload.

To create the animations, data pre-processing methodologies were applied to prepare the input data for the further workflow from various open datasets. It's notable that the

pitch map and the player's coordinate data should be matched, otherwise the visualization will be misplaced. Besides, the duration of the animation is worth to be careful because long-time of animation will not only make the users exhausted but also increases the total amount of data. Although every scene conveys rich information, it is too difficult to remember each frame of the animation if the length of the animation is too long. As a result, animations are better presented in less than a minute in duration.

Table 3 Software products and their manufacturers

<b>Solution</b>	<b>Strength</b>	<b>Application</b>
Python	In-depth analysis Tactics analysis	Pre-game and post-game analysis with Positioning, Convex hull, Voronoi diagram, and Marking
Web Technology	Interaction Customized setting	Cross-browser and cross-platform Real-time tactics arrangement
R	Low threshold Data standardization Save the workload	Entry-level football analyst practice Dealing with data from popular sports data providers

### Python

Rich libraries in Python developer community make it become a suitable solution for in-depth analysis. Bokeh is a good example. With Bokeh library, this thesis has four kinds of animations. Over the time slots, every scene of the pitch map is altered according to the specific players' and ball's location at that moment. Unfortunately, Bokeh library has a significant drawback. The standard export format is SVG and PNG. Thus, making a sharable animation outside the Python environment is not easy work. Furthermore, it's hard to execute this application outside Python environment.

Combined with GIS analysis, such as Convex hull, Voronoi diagram, and Shortest distance marking, the animations bring the viewer lots of new insights. For example, the players in the attacking team tend to pull apart the formation to make use of the space for open football, while the defending team always has a compact formation. Thus, the coverage area of convex hull from the attacking team is always bigger than the defending team. Voronoi diagram is particularly used for predicting or simulating interactions between neighbouring structures (see Fig. 30). Players tend to cluster together and compete around the football. The key value of Voronoi diagram is to clarify the responsibility and collaboration of defender. Voronoi diagram can help the coach to train the players on how to build a completed formation.

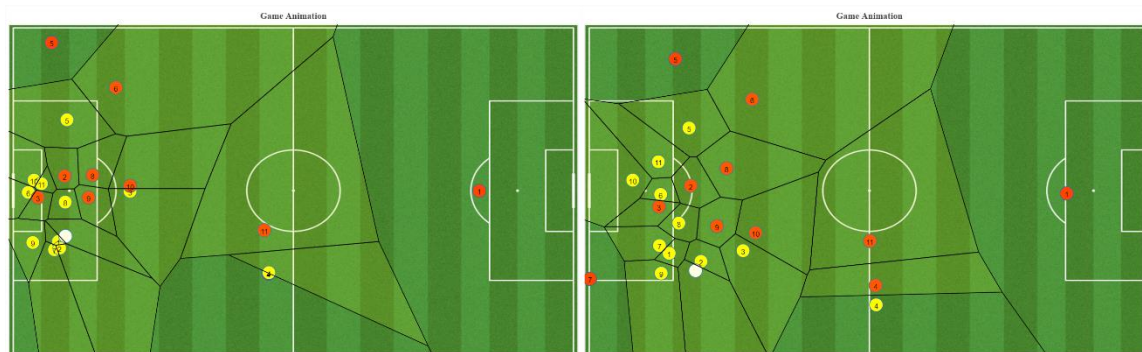


Figure 30 Animation of Voronoi diagram created by Python

## Web Technology

HTML5 is used for animation or interactivity, with the supplement of JavaScript and CSS. The whole solution is delivered as web interfaces structured in HTML5, styled with CSS, and its behaviour controlled by JavaScript. JavaScript contains lots of useful libraries such as jQuery. jQuery is a fast, small, and feature-rich library with a combination of versatility and extensibility.

The main advantage of this web solution is interaction and convenience. The target user group of this solution is the coach team. The coach prefers to use the vertical pitch map to emphasize the attack and defence direction, which is the reason that this solution is designed in vertical. 2.5D effect is also more real as a coach vision.

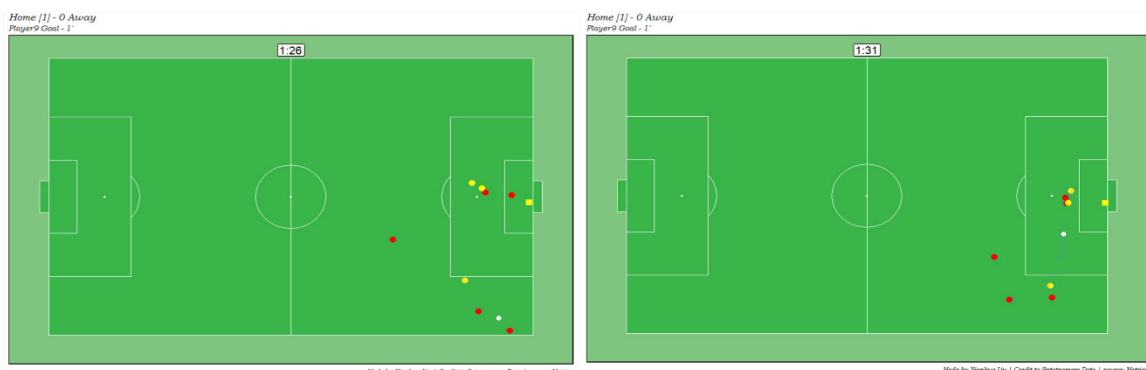
The main interface has two different layers: an object animation layer and a data input layer. Using the transformed trajectories, the movement of the individual players within the group can be presented quickly (see Fig. 31). Notably, to reduce the display clutter, users should not input too many trajectories at one time. One or a few trajectories should be input first and followed by more trajectories later.



**Figure 31** Animation of players movement created by Web Technology

## R

In this thesis, five packages are used to create the animation in R environment. The most important package among them is ggsoccer. ggsoccer is a package specialised in football visualization, which is implemented as a layer of an open-source data visualization package called ggplot. The biggest advantage of R solution is the standardization. ggsoccer is designed to work smoothly with a variety of popular sports data providers, which is easier to visualize the data from popular providers. A set of functions are provided by ggsoccer for elegantly displaying and exploring soccer event data, which makes it easy to plot data on a pitch. Besides, ggsoccer only has five indexes. The calling process of the package is not complicated. Although the customization is sacrificed, the learning threshold and workload is much less than others. Therefore, R is a good solution for beginner of football analysts, especially when they are using the football data from the popular business data provider.



**Figure 32** Animation of goal event created by R

## 7 DISCUSSION

By applying the developed workflows presented in this thesis, a set of visualizations and animations are created by successfully utilizing the potential of rich statistic, spatial, and spatio-temporal data. The presented contents can be regarded as a first approach toward the framework of static football visualizations and the application of football animations. However, a critical examination of the conducted work in this chapter reveals several weaknesses and limitations in the presented results. Although some progress has been made so far there is still a lot of work ahead.

### **Data acquisition**

Data is the most important factor in football visualization. In this thesis, all the data is collected by open data source instead of collecting raw data from hardware. Without the support of GPS device, the workflow is not completed from raw data collection. A future developing work of this thesis is connecting the hardware to the solution, which supports the automated visualization from the data collected by hardware.

A football game generates much more events and it is very important and interesting to take into account the context in which those events were generated. The anonymous open data may result in some loss of key information, compared to the raw data collected on a pitch. To collect the tracking data, the optical tracking solution is popular in elite clubs. But this solution has a budget requirement, which is a big challenge and high cost for small football clubs and personal researchers. GPS technology is utilised in some training sessions. But the usage under formal match is strict. In addition to automatically recorded tracking data, annotating the happened events manually with video is also possible. However, the manual workflow increases the uncertainty of the application, which requires assessment of data quality (Gennady Andrienko, Andrienko, & Fuchs, 2016).

### **Data pre-processing**

There is not a specific Football GIS software that existed now. Most of the football analysis software could not support the geo data format and sophisticated geospatial analysis. In this thesis, a bit of pre-processing work is necessary before it can be imported for visualization. This thesis aims at visualization. Choosing the qualified data from any potential data source takes precedence over working on the same dataset. Thus, all the open data from various providers are considered. The data format and type can vary from one to another, which is a challenge to unified and create an automated workflow. The pre-processing work is inevitable without any previous data preparation. Similarly, the background pitch map in this thesis is not unified because some packages and libraries have default pitch setting, which is not supported to change the parameters.

Moreover, the local coordinate system can vary from one provider to another. That's why the coordinate system for the visualization work in thesis is not unified. These circumstances must be considered when transforming the recorded data into pre-defined reference system in the future. Conventional football-specific geo data must be transformed from a Cartesian to a projected coordinate system prior to its implementation.

### **Data volume**

Besides, a huge amount of data for every game is common in football. More than 60 categories, about 140,000 positions for the ball and each player are recorded with spatio-temporal information. Considering all 22 players, the ball as well as the referee team, if applicable, about 3.5 million tracking data points are being captured during one game. Under these circumstances, real-time analysis is hard to implement during the game.



The vast amount of data leads to XML file data sizes of about 50 MB, which should be processed by some specific XML editor. This is particular the case if tracking and event data are combined and analysed (Kotzbek & Kainz, 2015b). In this thesis, web technology animated solution is sensitive to the data volume. If it is too large, the browser might be crashed.

### **Analysis**

A football game generates much more events than the visualized events in this thesis. It is very important to take into account the context in which those events were generated. Most publicly available football statistics are limited to aggregated data of regular events. When assessing performance, this simple aggregation, without any context, can be misleading (Kaggle, 2022). The effect of different events cannot be ascribed to the same impact in different spatio-temporal condition. Thus, the weighting of essential indicators is also a challenge. The analysis result in this thesis is based on the data, visualization, and animation. But this is far more than enough to know about a football match. Watching the full game itself should always be the first step to understand a match. Some important moment or turning point of a match can only be captured when the analyst watches the game first.

### **Application and verification**

To improve the user experience, some functions can be integrated. For example, the interaction methodology of a web animated solution can be chosen between the data input, and the drag and drop. For data input, it's easier to connect with the tracking hardware and transfer the data. However, the input time might be longer. For the drag and drop, it's easy for the coach to understand and adjust. However, the data is hard to transfer without access to the backend. Besides, the GIS analysis functions, such as Convex hull, Voronoi diagram, Buffer analysis, can also be integrated into web animation.

In fact, a good method to evaluate and verify the user experience is the user testing. Sending the survey to some specific groups like coach, football player, football analysis, and fans, to collect their feedback. With the results of the survey, the interface and interaction experience can better improved.

## 8 CONCLUSION

The aim of this thesis is to create a systematic framework of visualizations and explore different efficient workflow to create football animation. Football visualization becomes popular in recent years. However, the framework and workflow of static football visualization are still unclear.

Creating a framework of static football visualization is the first goal. Three groups of visualization works are created and the applied workflow, encountered issues, and the user scenario, are documented. In the course of this work, the visualization of football data is demonstrated and then analysed in detail. By taking the statistic character of football games adequately into account, the first part of the content establishes a systematic concept for all the football visualization charts, diagrams, statistic graphs, and plots. Next, examples of spatial information are created to demonstrate how visualization work highlight the players' location. The range of possibilities to apply GIS methods in football visualization and animation is examined, including Convex Hull, Voronoi diagram, and Buffer analysis. In addition, spatio-temporal information is analysed to emphasize the players' performance in a period. Heat maps and passing network maps are created to clarify the players' movement and passing events.

Static visualization methods are hard to deal with the increasing volumes of football data. The limitation of static football visualization comes with more challenges in visualizing dynamic football data. Thus, football animation is developing in recent years for better visualization. Exploring the range of possibilities and creating workflow of different programming languages is the second goal of this thesis.

The essential steps to animate football data through different technologies are described and important findings are shared. The solutions built by Python, Web technologies, and R are introduced in detail. Each of these animations has its style and special features to show its strength and application. With the rich libraries, Python is good at in-depth analysis, which is better used in pre-game and post-game analysis. Web Technologies can better reach the interaction effect, which can be a helpful tool for the coach in real-time analysis. With the standardization and low learning threshold, R has a better usage for entry-level analysis practice. Each animation constitutes a unique target group and user scenario. Some other solutions built by Tableau, ArcGIS, Adobe After Effects, and QGIS are also evaluated. The limitations concerning these methods are also documented. Coupling these football animations together with the toolset of different software can result in an advanced and intuitive medium, that allows football coaches and players to easily perceive and understand the depicted phenomena. The result of this process is animations based on different solutions, which were exported as GIF.

Overall, the results of the work will allow football stakeholders to learn more about the possibilities and applications of visualization and animation. The theoretical output of this thesis draws a framework of football visualization concept to what methodology can be used for different purposes. In addition, the practically created set of animations can demonstrate the value to different users and can give the viewer an idea of what information can be delivered with the utilized tools and the established workflow.

## REFERENCES AND INFORMATION SOURCES

- Albinsson, P.-A., & Andersson, D. (2008). Extending the Attribute Explorer to Support Professional Team-Sport Analysis. *7(2)*, 163-169. doi:10.1057/palgrave.ivs.9500178
- Andrienko, G., Andrienko, N., Anzer, G., Bauer, P., Budziak, G., Fuchs, G., . . . Wrobel, S. (2021). Constructing Spaces and Times for Tactical Analysis in Football. *IEEE Transactions on Visualization and Computer Graphics*, *27(4)*, 2280-2297. doi: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. doi:10.1007/s10618-017-0513-2
- Andrienko, G., Andrienko, N., & Fuchs, G. (2016). Understanding movement data quality. *Journal of Location Based Services*, *10(1)*, 31-46. doi:10.1080/17489725.2016.1169322
- Andrienko, N., Andrienko, G., Barrett, L., Dostie, M., & Henzi, P. (2013). Space Transformation for Understanding Group Movement. *IEEE Transactions on Visualization and Computer Graphics*, *19*, 2169-2178. doi:10.1109/TVCG.2013.193
- Andrienko, N., Andrienko, G., Pelekis, N., & Spaccapietra, S. (2008). Basic Concepts of Movement Data. In F. Giannotti & D. Pedreschi (Eds.), *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery* (pp. 15-38). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Andrienko, N. V., & Andrienko, G. L. (2005). *Exploratory analysis of spatial and temporal data - a systematic approach*.
- Animating goals in R using ggsoccer and gganimate* [online]. 2020-2022 [cit. 2022-04-28]. Tom MacNamara. Available online: <<https://statnamara.wordpress.com/2020/05/09/animating-goals-in-r-using-ggsoccer-and-gganimate/>>.
- Animating Time Series Data* [online]. 2022 [cit. 2022-04-30]. QGIS Tutorials and Tips. Available online: <[https://www.qgistutorials.com/en/docs/3/animating\\_time\\_series.html](https://www.qgistutorials.com/en/docs/3/animating_time_series.html)>.
- ArcGIS Maps for Adobe Creative Cloud* [online]. 2022 [cit. 2022-04-30]. Esri. Available online: <<https://www.esri.com/en-us/arcgis/products/maps-for-adobecc/overview>>.
- Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data Visualization and Evaluation. *2013(139)*, 7-32. doi:<https://doi.org/10.1002/ev.20065>
- Bartels, P., De Buyser, T., & Van Ussel, J. (2013). A Football Data Visualisation: The Belgian First Division. *Department of Computer Science, KU Leuven, Belgium*, 1-8.
- Barvir, R., & Vit, V. (2021). Graphic Map Load Measuring Tool – development and verification. *International Journal of Cartography*, *7(3)*, 285-303. doi:10.1080/23729333.2021.1972907
- Basole, R. C., & Saupe, D. (2016). Sports Data Visualization [Guest editors' introduction]. *IEEE Computer Graphics Applications*, *36*, 24-26.
- Beernaerts, J., De Baets, B., Lenoir, M., & Van de Weghe, N. (2020). Spatial movement pattern recognition in soccer based on relative player movements. *PloS one*, *15(1)*, e0227746. doi:10.1371/journal.pone.0227746
- Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2014). *Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data*. Paper presented at the 2014 IEEE International Conference on Data Mining.
- Bojinov, I. I., & Bornn, L. (2016). *The Pressing Game : Optimal Defensive Disruption in Soccer*.
- bokeh* [online]. 2022-2022 [cit. 2022-04-27]. James Raphael Tiovalen. Available online: <<https://github.com/bokeh/bokeh/blob/branch-3.0/examples/app/gapminder/main.py>>.

- Bradley, P., Ju, W., Laws, A., Gómez-Díaz, A., García, A., & Evans, M. (2020). FOOTBALL VISUALISATION: CAPTURING CHAOS AND CULTIVATING CONTEXT. *football medicine & performance*.
- Bradley, P., O'Donoghue, P., Wooster, B., & Tordoff, P. (2007). The reliability of ProZone MatchViewer: a video-based technical performance analysis system. *International Journal of Performance Analysis in Sport*, 7(3), 117-129. doi:10.1080/24748668.2007.11868415
- Bransen, L., Haaren, J. V., & Velden, M. v. d. (2019). Measuring soccer players' contributions to chance creation by valuing their passes. *Journal of Quantitative Analysis in Sports*, 15(2), 97-116. doi:doi:10.1515/jqas-2018-0020
- Buldú, J. M., Busquets, J., Echegoyen, I., & Seirullo, F. (2019). Defining a historic football team: Using Network Science to analyze Guardiola's F.C. Barcelona. *Scientific Reports*, 9(1), 13602. doi:10.1038/s41598-019-49969-2
- Caetano, F. G., Barbon Junior, S., Torres, R. d. S., Cunha, S. A., Ruffino, P. R. C., Martins, L. E. B., & Moura, F. A. (2021). Football player dominant region determined by a novel model based on instantaneous kinematics variables. *Scientific Reports*, 11(1), 18209. doi:10.1038/s41598-021-97537-4
- Carling, C., Williams, A. M., & Reilly, T. (2006). Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance. *Journal of sports science & medicine*, 5(1), 171-171.
- Cintia, P., Rinzivillo, S., & Pappalardo, L. (2015). *A network-based approach to evaluate the performance of football teams*. Paper presented at the Machine learning and data mining for sports analytics workshop, Porto, Portugal.
- Cotta, C., Mora, A., Molina, C., & Merelo Guervós, J. (2011). FIFA World Cup 2010: A Network Analysis of the Champion Team Play. *CoRR*, abs/1108.0261. doi:10.1007/s11424-013-2291-2
- Copernicus Services [online]. 2012–2022 [cit. 2022-03-24]. Copernicus Earth observation digital services. Available online: < <https://www.copernicus.eu/en/copernicus-services>>.
- Csanalosi, G., Dobreff, G., Pasic, A., Molnar, M., & Toka, L. (2020). *Low-Cost Optical Tracking of Soccer Players*. Paper presented at the Machine Learning and Data Mining for Sports Analytics, Cham.
- D'Orazio, T., & Leo, M. (2010). A review of vision-based systems for soccer video analysis. *Pattern Recognition*, 43(8), 2911-2926. doi:<https://doi.org/10.1016/j.patcog.2010.03.009>
- Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., . . . Weibel, R. (2015). Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*, 3(1), 5. doi:10.1186/s40462-015-0032-y
- Dickson, P. (1997). *The Joy of Keeping Score: How Scoring the Game Has Influenced and Enhanced the History of Baseball*. Harcourt Brace.
- Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Towards a Taxonomy of Movement Patterns. *Information Visualization*, 7, 240-252. doi:10.1057/palgrave.ivs.9500182
- Dransch, D. (2014). *Computer-Animation in der Kartographie: Theorie und Praxis*: Springer Berlin Heidelberg.
- Du, M., & Yuan, X. (2021). A survey of competitive sports data visualization and visual analysis. *Journal of Visualization*, 24(1), 47-67. doi:10.1007/s12650-020-00687-2
- Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). Quantifying the Performance of Individual Players in a Team Activity. *PloS one*, 5(6), e10937. doi:10.1371/journal.pone.0010937
- Fernández, J., & Bornn, L. (2018). *Wide Open Spaces: A statistical technique for measuring space creation in professional soccer*. Paper presented at the MIT Sloan Sports Analytics Conference.

- Fonseca, S., Milho, J., Travassos, B., Araujo, D., & Lopes, A. (2013). Measuring spatial interaction behavior in team sports using superimposed Voronoi diagrams. *International Journal of Performance Analysis in Sport*, 13, 179-189.
- football-crunching* [online]. 2019-2021 [cit. 2022-04-20]. Ricardo Tavares. Available online: <<https://github.com/rjtavares/football-crunching>>.
- Football Events* [online]. 2017 [cit. 2022-04-06]. Kaggle. Available online: <<https://www.kaggle.com/datasets/secareanualin/football-events>>.
- Football Storm* [online]. 2022 [cit. 2022-04-29]. Alexander Varlamov. Available online: <<https://public.tableau.com/app/profile/alexandervar>>.
- Frentzos, E., Theodoridis, Y., & N. Papadopoulos, A. (2009). Spatio-Temporal Trajectories. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 2742-2746). Boston, MA: Springer US.
- Fujimura, A., & Sugihara, K. (2005). Geometric analysis and quantitative evaluation of sport teamwork. 36(6), 49-58. doi:<https://doi.org/10.1002/scj.20254>
- Game-Animation* [online]. 2019-2021 [cit. 2022-04-26]. Samira Kumar. Available online: <<https://github.com/samirak93/Game-Animation>>.
- Gennady Andrienko, Natalia Andrienko, Peter Bak, Daniel Keim, Stefan Wrobel, & Heidelberg, B. (2013). Visual analytics of movement. *Annals of GIS*, 21(1), 91-92. doi:10.1080/19475683.2015.992828
- gganimate* [online]. 2022-2022 [cit. 2022-04-28]. Thomas Lin Pedersen. Available online: <<https://gganimate.com>>.
- ggplot2* [online]. 2022-2022 [cit. 2022-04-28]. Hadley Wickham. Available online: <<https://ggplot2.tidyverse.org>>.
- ggsoccer* [online]. 2022-2022 [cit. 2022-04-28]. Ben Torvaney. Available online: <<https://github.com/Torvaney/ggsoccer>>.
- Gonçalves, B., Coutinho, D., Exel, J., Travassos, B., Lago, C., & Sampaio, J. (2019). Extracting spatial-temporal features that describe a team match demands when considering the effects of the quality of opposition in elite football. *PloS one*, 14(8), e0221368-e0221368. doi:10.1371/journal.pone.0221368
- Gould, P., & Gatrell, A. (1979). A structural analysis of a game: The Liverpool v Manchester united cup final of 1977. *Social Networks*, 2(3), 253-273. doi:[https://doi.org/10.1016/0378-8733\(79\)90017-0](https://doi.org/10.1016/0378-8733(79)90017-0)
- Gudmundsson, J., & Horton, M. (2017). Spatio-Temporal Analysis of Team Sports -- A Survey. *ACM Computing Surveys*, 50(2). doi:<https://doi.org/10.1145/3054132>
- Gudmundsson, J., Laube, P., & Wolle, T. (2012). Computational Movement Analysis. In *Springer Handbook of Geographic Information*, 725-741.
- Gudmundsson, J., & Wolle, T. (2010). *Towards automated football analysis: Algorithms and data structures*. Paper presented at the Proc. 10th Australasian Conf. on mathematics and computers in sport.
- Gudmundsson, J., & Wolle, T. (2014). Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, 47, 16-27. doi:<https://doi.org/10.1016/j.compenvurbsys.2013.09.004>
- Harrie, L., & Weibel, R. (2007). Chapter 4 - Modelling the Overall Process of Generalisation. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski (Eds.), *Generalisation of Geographic Information* (pp. 67-87). Amsterdam: Elsevier Science B.V.
- Harrower, M., & Fabrikant, S. (2008). The Role of Map Animation for Geographic Visualization. *Geographic Visualization: Concepts, Tools and Applications*. doi:10.5167/uzh-8979
- Herold, M., Kempe, M., Bauer, P., & Meyer, T. (2021). Attacking Key Performance Indicators in Soccer: Current Practice and Perceptions from the Elite to Youth Academy Level. *Journal of sports science & medicine*, 20(1), 158-169. doi:10.52082/jssm.2021.158
- How to Draw a Football Pitch* [online]. 2020 [cit. 2022-04-20]. Peter McKeever. Available online: <<http://petermckeever.com/2020/10/how-to-draw-a-football-pitch>>.
- How to Make Animated Maps with ArcGIS Maps for Adobe and Adobe After Effects* [online]. 2020 [cit. 2022-04-29]. Elbert Leiva, Rey Umali. Available online: <

<https://www.esri.com/arcgis-blog/products/maps-for-adobecc/mapping/how-to-make-animated-maps>>.

- Hughes, M., & Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of sports sciences*, 23, 509-514. doi:10.1080/02640410410001716779
- Janetzko, H., Sacha, D., Stein, M., Schreck, T., Keim, D. A., & Deussen, O. (2014). *Feature-driven visual analytics of soccer data*. Paper presented at the 2014 IEEE Conference on Visual Analytics Science and Technology (VAST).
- jQuery [online]. 2022-2022 [cit. 2022-04-27]. jQuery. Available online: <<https://jquery.com>>.
- Kang, C., Hwang, J., & Li, K. (2006). *Trajectory Analysis for Soccer Players*. Paper presented at the Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06).
- Kim, H.-C., Kwon, O., & Li, K.-J. (2011). *Spatial and spatiotemporal analysis of soccer*. Paper presented at the Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, Illinois. <https://doi.org/10.1145/2093973.2094029>
- Kim, S. (2004). Voronoi Analysis of a Soccer Game. *Nonlinear Analysis. Modelling and Control*, 3. doi:10.15388/NA.2004.9.3.15154
- Kotzбек, G., & Kainz, W. (2014). *Football Game Analysis: A New Application Area For Cartographers And GI-Scientists?* Paper presented at the 5th International Conference on Cartography and GIS.
- Kotzбек, G., & Kainz, W. (2015a). Das Runde muss ins GIS – Neue Wege im Bereich der Fußball-Spielanalyse. *GIS-Zeitschrift für Geoinformatik*, 2015, 117-124.
- Kotzбек, G., & Kainz, W. (2015b). *GIS-Based Football Game Analysis - A Brief Introduction to the Applied Data Base and a Guideline on How to Utilise It*. Paper presented at the 27th International Cartographic Conference, Rio de Janeiro / Brazil.
- Kotzбек, G., & Kainz, W. (2016). *Towards Automated GIS-based Analysis of Scoring Attempt Patterns in Association Football*. Paper presented at the 19th AGILE Conference on Geographic Information Science, Helsinki, Finland.
- Kotzбек, G., & Kainz, W. (2017). *Beyond the Pass Completion Rate: Towards Analysing the Players' Passing Quality in Association Football*.
- Kotzбек, G., & Kainz, W. (2018). *Dynamic Zoning in the Course of GIS-Based Football Game Analysis*.
- Lacome, M., Simpson, B., & Buchheit, M. (2018). Part 2: Increasing Coach 'Buy-in' with Good Data Visualisation. *Aspetar Sports Medicine Journal*, 7, 64-66.
- Li, C., & Zhao, Y. (2021). Comparison of Goal Scoring Patterns in "The Big Five" European Football Leagues. 11. doi:10.3389/fpsyg.2020.619304
- Liu, N., Zhuang, L., Nêtek, R., & Strobl, J. (2021). *The overview, outcome, and outlook of GIS-based football visualization*. Paper presented at the 13th GeoMundus Conference 2021, Lisboa, Portugal.
- Lucey, P., Bialkowski, A., Carr, P., Foote, E., & Matthews, I. (2012). *Characterizing Multi-Agent Team Behavior from Partial Team Tracings : Evidence from the English Premier League* (Vol. 2).
- Lucey, P., Bialkowski, A., Monfort, M., Carr, P., & Matthews, I. (2015). "Quality vs Quantity": Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data.
- Maćkowiak, S., Konieczny, J., Kurc, M., & Maćkowiak, P. (2010). *Football Player Detection in Video Broadcast*. Paper presented at the Computer Vision and Graphics, Berlin, Heidelberg.
- Memmert, D. (2019). Data analytics in football: positional data collection, modeling, and analysis. *Journal of Sport Management*, 33(574), 2019-0308.
- mplsoccer [online]. 2020–2022 [cit. 2022-04-06]. Andrew Rowlinson. Available online: <<https://github.com/andrewRowlinson/mplsoccer>>.
- Narizuka, T., Yamamoto, K., & Yamazaki, Y. (2013). Statistical properties of position-dependent ball-passing networks in football games. *Physica A: Statistical Mechanics and its Applications*, 412. doi:10.1016/j.physa.2014.06.037

- open-data* [online]. 2022 [cit. 2022-04-30]. StatsBomb. Available online: < <https://github.com/statsbomb/open-data>>.
- 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. doi:10.1038/s41597-019-0247-7
- Peña, J. L., & Touchette, H. (2012). A network theory analysis of football strategies. *arXiv: Combinatorics*.
- Peñas, C., & A, D. (2010). Ball Possession Strategies in Elite Soccer According to the Evolution of the Match-Score: the Influence of Situational Variables. *Journal of Human Kinetics*, 25, 93-100. doi:10.2478/v10078-010-0036-z
- Perin, C., Vuillemot, R., & Fekete, J.-D. (2013). *Real-Time Crowdsourcing of Detailed Soccer Data*.
- Perin, C., Vuillemot, R., & Fekete, J. (2013). SoccerStories: A Kick-off for Visual Soccer Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2506-2515. doi: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. 37(3), 663-686. doi:<https://doi.org/10.1111/cgf.13447>
- Pettersen, S. A., Johansen, H. D., Baptista, I. A. M., Halvorsen, P., & Johansen, D. (2018). Quantified Soccer Using Positional Data: A Case Study. *Frontiers in physiology*, 9, 866-866. doi:10.3389/fphys.2018.00866
- QSoccer* [online]. 2021-2022 [cit. 2022-04-29]. Oslandia. Available online: < <https://oslandia.gitlab.io/qgis/QSoccer/>>.
- R* [online]. 1991-2022 [cit. 2022-04-28]. R Foundation. Available online: < <https://www.r-project.org>>.
- Reep, C., & Benjamin, B. (1968). Skill and Chance in Association Football. *Journal of the Royal Statistical Society. Series A (General)*, 131(4), 581-585. doi:10.2307/2343726
- Rodrigues, P., Cardoso, P., & Rodrigues, J. M. F. (2013). *A Field, Tracking and Video Editor tool for a Football Resource Planner*. Paper presented at the 2013 8th Iberian Conference on Information Systems and Technologies (CISTI).
- RStudio* [online]. 2009-2022 [cit. 2022-04-28]. RStudio. Available online: < <https://www.rstudio.com>>.
- Rusu, A., Stoica, D., Burns, E., Hample, B., McGarry, K., & Russell, R. (2010). *Dynamic Visualizations for Soccer Statistical Analysis*. Paper presented at the 2010 14th International Conference Information Visualisation.
- Sacha, D., Al-Masoudi, F., Stein, M., Schreck, T., Keim, D., Andrienko, G., & Janetzko, H. (2017). Dynamic Visual Abstraction of Soccer Movement. *Computer Graphics Forum*, 36, 305-315. doi:10.1111/cgf.13189
- sample-data* [online]. 2020-2022 [cit. 2022-04-28]. Metrica Sports. Available online: < [https://github.com/metrica-sports/sample-data/tree/master/data/Sample\\_Game\\_1](https://github.com/metrica-sports/sample-data/tree/master/data/Sample_Game_1)>.
- Sarmiento, H., Marcelino, R., Campanico, J., Matos, N., & Leitão, J. (2014). Match analysis in football: a systematic review. *Journal of sports sciences*, 32, 1831-1843. doi:10.1080/02640414.2014.898852
- Schäfer, N. O. (2021). *Map animation of shipping traffic in Adobe After Effects from GIS data*. Palacký University Olomouc, Olomouc.
- Sester, M., Feuerhake, U., Kuntsch, C., & Zourlidou, S. (2015). Interpretation of Moving Point Trajectories. *Photogrammetric Week*(15), 329-341.
- Shao, L., Sacha, D., Neldner, B., Stein, M., & Schreck, T. (2016). Visual-Interactive Search for Soccer Trajectories to Identify Interesting Game Situations. *Electronic Imaging*, 2016, 1-10. doi:10.2352/ISSN.2470-1173.2016.1.VDA-510
- soccerplots* [online]. 2020-2022 [cit. 2022-04-06]. Anmol Durgapal. Available online: <<https://github.com/Slothfulwave612/soccerplots>>.

- Stein, M., Häußler, J., Jäckle, D., Janetzko, H., Schreck, T., & Keim, D. A. (2015). Visual Soccer Analytics: Understanding the Characteristics of Collective Team Movement Based on Feature-Driven Analysis and Abstraction. *4*(4), 2159-2184.
- Stein, M., Janetzko, H., Breitzkreutz, T., Seebacher, D., Schreck, T., Grossniklaus, M., . . . Keim, D. A. (2016). Director's Cut: Analysis and Annotation of Soccer Matches. *IEEE Computer Graphics and Applications*, *36*(5), 50-60. doi:10.1109/MCG.2016.102
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *12*(2), 257-285. doi:[https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)
- Taki, T., Hasegawa, J.-i., & Fukumura, T. (1996). Development of motion analysis system for quantitative evaluation of teamwork in soccer games. *Proceedings of 3rd IEEE International Conference on Image Processing*, *3*, 815-818 vol.813.
- Velden, M. v. d., Haaren, J. V., Franses, P. H., & Hoeve, L. T. v. (2017). *Valuing passes in football using ball event data*.
- Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites. *Journal für die reine und angewandte Mathematik*, 97 - 102.
- Wasserman, S. F. K. (1994). *Social network analysis : methods and applications*.
- Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology*, *13*(4), e1002128. doi:10.1371/journal.pbio.1002128
- Weissgerber, T. L., Winham, S. J., Heinzen, E. P., Milin-Lazovic, J. S., Garcia-Valencia, O., Bukumiric, Z., . . . Milic, N. M. (2019). Reveal, Don't Conceal: Transforming Data Visualization to Improve Transparency. *Circulation*, *140*(18), 1506-1518. doi:10.1161/circulationaha.118.037777
- Wikstrand, G., & Eriksson, S. (2002). *Football animations for mobile phones*. Paper presented at the Proceedings of the second Nordic conference on Human-computer interaction, Aarhus, Denmark. <https://doi.org/10.1145/572020.572059>
- Wirfs-Brock, A., & Eich, B. (2020). JavaScript: the first 20 years. *4*(HOPL %J Proc. ACM Program. Lang.), Article 77. doi:10.1145/3386327
- Yamamoto, Y. (2009). Scale-free Property of the Passing Behaviour in a Team Sport. *International Journal of Sport and Health Science*, *7*, 86-95. doi:10.5432/ijshs.IJSHS20090014
- Yamamoto, Y., & Yokoyama, K. (2011). Common and Unique Network Dynamics in Football Games. *PloS one*, *6*(12), e29638. doi:10.1371/journal.pone.0029638



# **ATTACHMENTS**

# LIST OF ATTACHMENTS

## Free attachments

- Attachment 1 Poster
- Attachment 2 Flash drive

## Structure Flash Drive

### Folders:

- Animation
  - Python\_convex.gif
  - Python\_marking.gif
  - Python\_Positioning.gif
  - Python\_voronoi.gif
  - R.gif
  - web.gif
- Code\_Data
  - Bookman Old Style.ttf
  - Column\_chart.ipynb
  - Column\_chart\_input
    - ...
  - Heatmap.ipynb
  - Line\_plot.ipynb
  - Movement\_tracking\_map.ipynb
  - Movement\_tracking\_map\_input
    - ...
  - Passes\_event\_arrow\_map.ipynb
  - Passes\_event\_scatter\_map.ipynb
  - Passing\_network\_map.ipynb
  - Python\_Animation
    - ...
  - R\_Animation
    - ...
  - Pitch\_map.ipynb
  - Radar\_chart.ipynb
- Figure
- Poster
- Thesis
- Website\_Football
- Website\_Thesis